THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *27 Juin 2017* **par :**

Fanny Augeri

Principes de grandes déviations pour des modèles de matrices aléatoires

---

JURY

Florent Benaych-Georges	Université Paris Descartes	Examinateur
Charles Bordenave	Université Toulouse 3	Directeur de thèse
Djalil Chafaï	Université Paris-Dauphine	Examinateur
Michel Ledoux	Université Toulouse 3	Examinateur
Mylène Maïda	Université Lille 1	Rapportrice

École doctorale et spécialité :
MITT : Domaine Mathématiques : Mathématiques appliquées
Unité de Recherche :
Institut de Mathématiques de Toulouse (UMR 5219)
Directeur de Thèse :
Charles Bordenave
Rapporteurs :
Mylène Maïda et *Ofer Zeitouni*

Table of contents

1. Introduction

1.1 Motivations

L'étude des matrices aléatoires tire une de ses origines de la mécanique statistique des noyaux d'atomes lourds étudiés par Wigner dans les années 50. Il proposa l'idée que les niveaux d'énergies de ces noyaux auraient le même comportement statistique que les valeurs propres de matrices de l'ensemble orthogonal Gaussien (GOE). Depuis, l'essor des matrices aléatoires a dépassé largement le cadre de la physique mathématique et a suscité un intérêt grandissant dû à l'ubiquité dont a fait preuve ces objets. La théorie des matrices aléatoires a montré des connections avec des domaines aussi variés que les probabilités libres, les algèbres d'opérateurs, la théorie des graphes, la combinatoire, les modèles intégrables, la théorie analytique des nombres, les polynômes orthogonaux et les processus déterminantaux.

Une des questions qui traverse la théorie des matrices aléatoires est celle de l'*universalité*. Simultanément à la modélisation des niveaux d'énergie de noyaux d'atomes lourds par les valeurs propres des matrices du GOE, Wigner avança l'idée qu'étant donné un système satisfaisant une certaine symétrie, considérer un Hamiltonien "typique" dans la classe de symétrie de ce système suffit statistiquement à le décrire. Dyson [47] montra que les seuls groupes de symétrie possibles sont les groupes orthogonal, unitaire et symplectique, ce qui permit d'identifier trois classes d'universalité dont les prototypes sont les ensembles Gaussiens classiques vérifiant chacun une de ces symétries : l'ensemble orthogonal Gaussien (GOE), l'ensemble unitaire Gaussien (GUE), et l'ensemble Gaussien symplectique (GSE).

D'un point de vue mathématique, une des façons d'interpréter cette conjecture d'universalité est de déterminer quels comportements du spectre de matrices aléatoires dans une des classes de symétrie sont universels, au sens où ils ne dépendent pas de la loi des entrées de la matrice. Le comportement macroscopique du spectre est connu pour être universel, comme le montre le célèbre théorème de Wigner. L'universalité du comportement mésoscopique du spectre, des espacements entre les valeurs propres, de la délocalisation des vecteurs propres, a été montré récemment sous des hypothèses d'intégrabilité suffisante des entrées par Tao-Vu d'une part et Erdös-Schlein-Yau d'autre part.

Parmi les problématiques d'universalité, le comportement de répulsion entre les valeurs propres des matrices aléatoires, la délocalisation complète de ses vecteurs propres, est souvent opposé au comportement Poissonien des processus ponctuels et aux phénomènes de localisation. Les conjectures de Bohgias-Giannoni-Schmidt et de Berry-Tabor sur les valeurs propres du Laplacien d'un espace compact, prédisent qu'il existe une dichotomie entre les systèmes intégrables qui sont censés suivre le même comportement que celui du spectre des matrices aléatoires, et celui des systèmes intégrables qui devraient suivre les mêmes statistiques qu'un processus de Poisson.

1.2 Phénomène de répulsion des valeurs propres

Comme on vient de le mentionner, une des caractéristiques surprenante du spectre des matrices aléatoires est que les valeurs propres ont tendance à se repousser. Nous allons essayer de donner quelques arguments simples (de nature déterministe) permettant de voir ce phénomène.

Notons $\mathcal{H}_n^\beta$ l'espace des matrices symétriques de taille n si $\beta = 1$, et Hermitiennes si $\beta = 2$. Un théorème célèbre de Von Neumann et Wigner [101] dit que l'ensemble des matrices de $\mathcal{H}_n^\beta$ ayant au moins une valeur propre multiple est une sous-variété de codimension (réelle) 2 si $\beta = 1$ et 3 si $\beta = 2$, ce qui traduit une sorte de répulsion entre les valeurs propres. En particulier, une matrice aléatoire Hermitienne dont la loi est absolument continue par rapport à la mesure de Lebesgue sur $\mathcal{H}_n^\beta$, a presque sûrement un spectre simple.

A l'inverse, si X est une matrice Hermitienne à coefficients entiers, le discriminant

$$D = \prod_{i \neq j} (\lambda_i - \lambda_j),$$

où $\lambda_1, ..., \lambda_n$ sont les valeurs propres de X, est un nombre entier car D est une fonction polynomiale symétrique des λ_i qui s'exprime donc comme polynôme en les polynômes symétriques élémentaires. En particulier si $D \neq 0$, on a alors $D \geq 1$.

Un dernier exemple de répulsion est donné dans le cadre des matrices de Jacobi,

$$\begin{pmatrix} b_1 & a_1 & 0 & \cdots & 0 \\ a_1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & a_{n-1} \\ 0 & \cdots & 0 & a_{n-1} & b_n \end{pmatrix},$$

où les a_i sont des coefficients strictement positifs. On sait d'après [41, Chapitre 3 §3.1] que,

$$\prod_{i=1}^{n-1} a_i^{2(n-i)} = D \prod_{i=1}^{n} f_i^2(1),$$

où $f_1, ..., f_n$ est une famille orthonormée de vecteurs propres. On en déduit que pour $a_i = 1$, en utilisant l'inégalité aritmético-géometrique,

$$D^{\frac{1}{n(n-1)}} \geq n^{\frac{1}{n-1}} \geq 1,$$

ce qui donne encore une fois une indication de la répulsion entre les valeurs propres.

1.3 Comportement global

Dans la suite, on s'intéressera principalement aux matrices de Wigner. On dira qu'une matrice aléatoire Hermitienne X est une *matrice de Wigner* si $(X_{i,j})_{i<j}$ et $(X_{i,i})_i$ sont deux familles indépendantes, de variables indépendantes et identiquement distribuées (i.i.d) de loi ne dépendant pas de la dimension.

1.3.1 Théorème de Wigner

On associe à une matrice aléatoire X de taille $n \times n$ à coefficients complexes, une mesure de probabilité aléatoire sur $\mathbb{C}$, appelée *mesure spectrale empirique*, notée μ_X et définie par,

$$\mu_X = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i},$$

où $\lambda_1, ..., \lambda_n$ désignent les valeurs propres de X. On observe que si X est une matrice normale, on a pour $\alpha > 0$,

$$\mu_X(|x|^\alpha) = \frac{1}{n}\sum_{i=1}^{n} s_i^\alpha,$$

où $s_1, ..., s_n$ sont les valeurs singulières de X. Si $\alpha \leq 2$, d'après [104, Théorème 3.32], on a la comparaison suivante entre les entrées et les valeurs singulières,

$$\sum_{i=1}^{n} s_i^\alpha \leq \sum_{i,j} |X_{i,j}|^\alpha. \tag{1.1}$$

On en déduit que si les entrées de X ont des moments d'ordre α finis, on a

$$\mathbb{E}\mu_{X/n^{1/\alpha}}(|x|^\alpha) = O(1).$$

Ceci montre que la suite $\mathbb{E}\mu_{X/n^{1/\alpha}}$ est tendue pour la topologie faible. Dans le cas où X est une matrice de Wigner avec des entrées ayant un moment d'ordre 2 fini, la convergence de la mesure spectrale empirique est donnée par le célèbre théorème suivant [103], [3, Théorème 2.1.1].

1.3.1 Théorème de Wigner. *Soit X une matrice de Wigner telle que $\mathbb{E}|X_{1,2} - \mathbb{E}X_{1,2}|^2 = 1$. Presque sûrement,*

$$\mu_{X/\sqrt{n}} \underset{n\to+\infty}{\rightsquigarrow} \mu_{sc},$$

où $\rightsquigarrow$ désigne la convergence pour la topologie faible et μ_{sc} est la loi du semi-cercle définie par,

$$\mu_{sc} = \mathbb{1}_{t\in[-2,2]}\frac{1}{2\pi}\sqrt{4-t^2}dt.$$

Le fait que l'hypothèse d'intégrabilité ne porte que sur les entrées hors-diagonales reflète le rôle négligeable que joue les entrées diagonales dans le comportement global du spectre, puisqu'elles constituent une petite proportion seulement des entrées. Ceci est essentiellement dû à l'inégalité d'Hoeffman-Wielandt [22, Théorème VI.4.1],

$$\mathcal{W}_2(\mu_A, \mu_B) \leq \frac{1}{n^{1/2}} ||A - B||_{\ell^2}, \tag{1.2}$$

où A, B sont deux matrices normales de taille n, et $\mathcal{W}_2$ est la distance L^2-Wasserstein définie par,

$$\mathcal{W}_2(\mu, \nu)^2 = \inf_{\pi} \int |x - y|^2 d\pi(x, y),$$

où l'infimum porte sur tous les couplages π entre μ et ν. (On utilise implicitement dans (1.2) le fait que l'infimum au-dessus est atteint pour des couplage π qui correspondent à des appareillements, grâce au théorème de Birkhoff [22, Théorème II.2.3]). L'asymptotique de la mesure spectrale empirique est stable sous des perturbations de rang fini, grâce aux inégalités d'entrelacement [22, Théorème III.2.1], [8, Théorème A.43],

$$d_{KS}(\mu_A, \mu_B) \leq \frac{1}{n} \mathrm{rank}(A - B), \tag{1.3}$$

où A et B sont deux matrices Hermitiennes et d_{KS} désigne la distance de Kolmogorov-Smirnov. Cette stabilité implique que l'asymptotique de la mesure spectrale empirique ne dépend pas de la moyenne des entrées hors-diagonale. Ceci explique pourquoi seule une hypothèse de normalisation de la variance des entrées hors-diagonales est pertinente.

1.3.2 Méthode des moments

La preuve originelle de Wigner repose sur la *méthode des moments*. Comme la mesure limite est à support compacte, elle est caractérisée par ses moments, ce qui permet de se restreindre à montrer seulement la convergence des moments de la mesure spectrale empirique, vers ceux de la loi du semi-cercle,

$$\frac{1}{n} \mathrm{tr}(X/\sqrt{n})^p \underset{n \to +\infty}{\longrightarrow} \mu_{sc}(x^p).$$

Par symétrie, $\mu_{sc}(x^{2p+1}) = 0$, tandis que les moments pairs $\mu_{sc}(x^{2p})$ du semi-cercle sont donnés par les *nombres de Catalan* C_p. Ces nombres comptent de nombreux objets combinatoires (66 d'après [92, Exercice 6.19]!) qui ont une structure récursive d'arbre. Ici, les nombres de Catalan C_p apparaissent naturellement comme le nombre d'arbres planaires enracinés ayant $p+1$ sommets [92, Exercice 6.19, e]. Des arguments de troncation et de recentrage (voir [8, Théorème 2.5]) permettent de se ramener au cas où les entrées diagonales sont nulles, les entrées hors-diagonales sont centrées et bornées. En utilisant un argument de concentration, la preuve du Théorème 1.3.1 se réduit à montrer,

$$\frac{1}{n} \mathbb{E} \mathrm{tr}(X/\sqrt{n})^p \underset{n \to +\infty}{\longrightarrow} \begin{cases} C_{p/2} & \text{si } p \in 2\mathbb{Z}, \\ 0 & \text{sinon.} \end{cases}$$

En analysant le développement de la trace, et en triant les mots qui apparaissent suivant le graphe qui les supporte, on en déduit à cause de la normalisation en $n^{1+p/2}$ que seuls les mots supportés par des arbres planaires à $p/2+1$ sommets peuvent survivre à la limite, ce qui permet de voir apparaître les nombres de Catalan.

Cette preuve due à Wigner a ouvert la voie à une approche combinatoire de l'étude du spectre des matrices aléatoires qui a trouvé de nombreux développements sous le nom de la “méthode des traces de Wigner” que l'on essaiera d'illustrer dans la suite.

1.3.3 Équation de point fixe de la résolvante

Une autre approche plus analytique au théorème de Wigner, due à Marchenko et Pastur, consiste à étudier la convergence des transformées de Stieltjes. Pour une mesure de probabilité μ sur $\mathbb{R}$, on définit sa *transformée de Stieltjes* par,

$$\forall x \in \mathbb{C}^+,\ g_\mu(z) = \int \frac{d\mu(x)}{z-x},$$

où $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$. Cette formule définit une fonction holomorphe sur $\mathbb{C}_+$ qui caractérise la loi μ. En effet, $\Im g_\mu(.+i\varepsilon)$ correspond à la densité de μ convolée par $\eta(\varepsilon^{-1}.)$ où η est la loi de Cauchy. Si μ et ν ont même transformée de Stieltjes, on en déduit que

$$\mu * \eta(\varepsilon^{-1}.) = \nu * \eta(\varepsilon^{-1}.),$$

ce qui implique, puisque $\eta(\varepsilon^{-1}.)$ converge faiblement vers δ_0 quand $\varepsilon \to 0$, que $\mu = \nu$. A cause de l'analyticité, la convergence faible d'une suite de mesures de probabilité μ_n vers μ est équivalente à la convergence de leurs transformées de Stieltjes sur un compact contenant un point d'accumulation (voir [3, Théorème 2.4.4]).

Ceci permet de définir la distance suivante sur l'espace des mesures de probabilités sur $\mathbb{R}$, noté $\mathcal{P}(\mathbb{R})$, muni de sa tribu Borélienne,

$$d(\mu,\nu) = \sup\big\{|g_\mu(z) - g_\nu(z)| : z \in \mathcal{K}\big\}, \tag{1.4}$$

où $\mathcal{K}$ désigne un ensemble compact de $\mathbb{C}^+$ contenant un point d'accumulation.

En conséquence de l'équation de récursion vérifiée par les nombres de Catalan, la transformée de Stieltjes de la loi du semi-cercle est l'unique solution qui s'annule à l'infini, de l'équation de point fixe suivante (voir [3, Lemme 2.1.3]),

$$\forall z \in \mathbb{C}^+,\ g_{\mu_{\mathrm{sc}}}(z) + \frac{1}{g_{\mu_{\mathrm{sc}}}(z)} = z. \tag{1.5}$$

Le point de départ de cette approche analytique, que l'on retrouve dans beaucoup de situations en théorie spectrale, notamment dans les problèmes de stabilité du spectre absolument continu (voir par exemple [65], [1]), est de chercher une équation “auto-similaire” vérifiée par la résolvante $G = (z - X/\sqrt{n})^{-1}$ définie pour $z \in \mathbb{C}^+$. Celle-ci est donnée par la formule du complément de Schur (voir [3, Lemma 2.4.6]),

$$\frac{1}{G_{i,i}} = z - X_{i,i}/\sqrt{n} - x_i^* G^{(i)} x_i,$$

où x_i désigne le $i^{\text{ème}}$ vecteur colonne de X dont on a retiré le $i^{\text{ème}}$ coefficient et $G^{(i)}$ la résolvante de la matrice $X/\sqrt{n}$ dont on a enlevé les $i^{\text{ème}}$ ligne et colonne. Supposons X centrée. Informellement, en utilisant l'indépendance de $G^{(i)}$ et x_i, et des arguments de concentration, on en déduit,

$$\frac{1}{G_{i,i}} \simeq z - \frac{1}{n}\mathrm{tr}G^{(i)}.$$

Comme on a retiré une proportion négligeable d'entrées de X, on peut voir grâce à l'inégalité d'Hoeffman-Wielandt (1.2), que

$$\frac{1}{G_{i,i}} \simeq z - \frac{1}{n}\mathrm{tr}G,$$

ce qui implique en faisant la moyenne sur i,

$$1 \simeq \frac{z}{n}\mathrm{tr}G - \left(\frac{1}{n}\mathrm{tr}G\right)^2.$$

On en déduit par un argument de compacité, la convergence de la transformée de Stieltjes de $\mu_{X/\sqrt{n}}$ vers celle de μ_{sc}.

1.3.4 Matrices de Wigner à queues lourdes

Dans le cas où les entrées n'ont pas de moment d'ordre 2, la convergence de la mesure spectrale empirique est connue grâce au résultat de Ben Arous-Guionnet [17] pour des entrées suivant une loi α-stable pour $\alpha \in (0,2)$. L'argument présenté au début de la section 1.3.1 montre que dans ce cas, la suite $\mathbb{E}\mu_{X/N}$, où $N = n^{1/(\alpha-\varepsilon)}$, est tendue pour tout $\varepsilon > 0$, ce qui donne le bon ordre de grandeur de la normalisation nécessaire : la mesure spectrale empirique de X/a_n converge en probabilité, avec $a_n = n^{1/\alpha}L(n)$ où L est une fonction à variations lentes, vers une mesure μ_α qui ne dépend que de α, et qui a la particularité de ne pas être à support compact.

1.4 Convergence du support

La loi du semi-cercle étant à support compact, il est légitime de se demander, dans le cadre du théorème de Wigner, si les valeurs extrêmes du spectre d'une matrice de Wigner converge vers les bords du support de μ_{sc}. Dans la suite on note λ_X la plus grande valeur propre d'une matrice aléatoire Hermitienne X. On observe que l'application,

$$\mu \in \mathcal{P}(\mathbb{R}) \mapsto \sup \mathrm{supp}\mu,$$

est semi-continue inférieurement pour la topologie faible. Le théorème de Wigner implique donc,

$$\liminf_{n\to+\infty} \lambda_{X/\sqrt{n}} \geq 2 \quad \text{p.s.}$$

La convergence de $\lambda_{X/\sqrt{n}}$ vers 2 est liée à l'intégrabilité des entrées, car

$$\lambda_{X/\sqrt{n}} = \max_{||u||\leq 1} \langle u, Xu\rangle.$$

Le résultat suivant montre que sous l'hypothèse d'un moment d'ordre 4 fini des entrées hors-diagonales, on a bien la convergence de la plus grande valeur propre vers 2.

1.4.1 Théorème. *Soit $(X_{i,j})_{i\leq j}$ une famille de variables aléatoires indépendantes telle que $(X_{i,i})_{i\geq 1}$ sont i.i.d et $(X_{i,j})_{i<j}$ sont i.i.d. On forme la matrice de Wigner $X = (X_{i,j})_{1\leq i,j\leq n}$. On suppose $\mathbb{E}X_{1,2} = 0$, $\mathbb{E}(X_{1,1})_+^2 < +\infty$, $\mathbb{E}|X_{1,2}|^2 = 1$, et $\mathbb{E}|X_{1,2}|^4 < +\infty$. Presque sûrement,*

$$\lambda_{X/\sqrt{n}} \underset{n\to+\infty}{\longrightarrow} 2.$$

Le couplage choisi ici est crucial pour obtenir la convergence presque sûre de la plus grande valeur propre. Cette construction est seulement pratique techniquement, mais n'a pas de sens particulier d'un point de spectral pour les matrices de Wigner. On note que pour la convergence en probabilité, on peut affaiblir légèrement les hypothèses d'intégrabilité des entrées (voir [8, Théorème 5.3])

Par des arguments de troncation et recentrage usuels, on peut montrer qu'il suffit de prouver le résultat du Théorème 1.4.1 pour X ayant des entrées centrées et ayant des moments qui ne croissent pas trop vite. Une preuve de cet énoncé consiste à évaluer de grandes traces, car pour $k \gg \log n$,

$$|\lambda_{X/\sqrt{n}}| = \left(\frac{1}{n}\mathrm{tr}(X/\sqrt{n})^{2k})\right)^{1/2k}(1+o(1)).$$

L'analyse de ces grandes traces repose sur des méthodes combinatoires d'énumération de chemins dues à Füredi-Kolmós [50], qui ont été raffinées successivement par exemple par Vu [102], ou Sinaĭ et Soshnikov [90].

On peut voir facilement que l'hypothèse $\mathbb{E}(X_{1,1_+}) < +\infty$ est nécessaire, puisque si $\mathbb{E}(X_{1,1})_+^2 = +\infty$, et $t > 0$, alors on peut montrer que

$$\sum_{n=1}^{+\infty} \mathbb{P}((X_{n,n})_+ \geq t\sqrt{n}) = +\infty,$$

d'où par le Lemme de Borel-Cantelli,

$$\mathbb{P}(\limsup_{n\to+\infty} \lambda_{X/\sqrt{n}} \geq t) = 1.$$

Donc $\limsup \lambda_{X/\sqrt{n}} = +\infty$ p.s. Le moment d'ordre 4 des entrées hors-diagonales est aussi une condition nécessaire, par un résultat de Bai et Yin [10]. Celle-ci est un peu plus subtile, puisque la participation des entrées hors-diagonales dans les déviations de la plus grande valeur propre est un peu plus compliquée que celle des entrées diagonales. Ceci dit, par le même argument que ci-dessus, on peut voir que si $\mathbb{E}X_{1,2}^4 = +\infty$, on a presque sûrement

$$\limsup_{n\to+\infty} ||X/\sqrt{n}|| = +\infty.$$

En effet, $||X|| = \max_{||u||,||v||\leq 1} |\langle u, Xv\rangle|$ d'une part, et d'autre part on peut montrer que $\mathbb{E}X_{1,2}^4 = +\infty$ implique,

$$\sum_{n=1}^{+\infty} \mathbb{P}(\max_{j<n} |X_{n,j}| \geq t\sqrt{n}) = +\infty.$$

Ceci montre par le Lemme de Borel-Cantelli,

$$\mathbb{P}(\limsup_{n\to+\infty} ||X/\sqrt{n}|| \geq t) = 1.$$

Comme pour le comportement de la mesure spectrale empirique, la moyenne des entrées diagonales ne joue pas de rôle, puisqu'une perturbation diagonale revient à translater le spectre de $X/\sqrt{n}$ de l'ordre de $O(n^{-1/2})$. Par contre changer la moyenne des entrées hors diagonales a pour effet de perturber le bord du spectre. En effet, changer $\mathbb{E}\Re X_{1,2}$ revient à faire une perturbation de rang 1 de $X/\sqrt{n}$ de valeur propre non nulle d'ordre $O(n^{1/2})$, ce qui, à cause des inégalités de Weyl (voir [22, Corollaire III.2.2]),

$$\lambda_k(A) + \lambda_1(H) \leq \lambda_k(A+H) \leq \lambda_k(A) + \lambda_n(H), \tag{1.6}$$

où $\lambda_1(A) \leq .... \leq \lambda_n(A)$ sont les valeurs propres de A, fait diverger le bord (gauche ou droite suivant le signe de $\mathbb{E}\Re X_{1,2}$) du spectre. De la même façon, une perturbation de $\mathbb{E}\Im X_{1,2}$ correspond à une translation de $X/\sqrt{n}$ par une matrice de rayon spectral d'ordre $O(n^{1/2})$, puisque d'après [8, Lemme 2.7],

$$\sigma(H) = \left\{\cot\left(\pi\frac{2k-1}{2n}\right) : k = 1, ..., n\right\},$$

où $\sigma(H)$ désigne le spectre de la matrice H de taille $n \times n$,

$$H = \begin{pmatrix} 0 & i & \cdots & i \\ -i & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & i \\ -i & \cdots & -i & 0 \end{pmatrix}.$$

1.5 Modèles déformés

Une question fondamentale en théorie spectrale est celle de la stabilité du spectre d'un opérateur (typiquement du Laplacien dans un contexte continu ou discret) sous des perturbations. Dans ce paragraphe, nous allons voir quelles réponses on peut apporter dans le cadre des matrices de Wigner. La situation typique est la suivante :

$$W_n = X/\sqrt{n} + A_n,$$

où X est une matrice de Wigner et A_n est une suite de matrices Hermitiennes déterministes de rayon spectral uniformément borné telle que,

$$\mu_{A_n} \underset{n\to+\infty}{\rightsquigarrow} \mu,$$

où μ est une mesure de probabilité à support compact. Quel est alors le comportement asymptotique de la mesure spectrale empirique de la matrice déformée W_n? Celui des bords du spectre de W_n? Ces questions trouvent dans les probabilités libres un formalisme adéquat. Concernant le comportement global du spectre, si X a des moments finis à tout ordre, Speicher [91] a montré que presque sûrement,

$$\mu_{W_n} \underset{n\to+\infty}{\rightsquigarrow} \mu_{sc} \boxplus \mu,$$

où $\boxplus$ désigne la convolution libre entre mesures de probabilité. Celle-ci peut se définir par sa R-tranformée. Pour une mesure de probabilité μ sur $\mathbb{R}$, on définit sur un voisinage de ∞ de $\mathbb{C}^+$, la fonction,

$$R_\mu(z) = g_\mu^{-1}(z) - \frac{1}{z}.$$

Puisque la dérivée de g_μ en ∞ est égale à 1, g_μ admet un inverse sur un certain voisinage de l'infini, ce qui donne bien un sens à R_μ, qui est donc une fonction analytique qui caractérise μ. La R-transformée de la convolution libre $\mu \boxplus \nu$ s'écrit,

$$R_{\mu \boxplus \nu} = R_\mu + R_\nu,$$

où l'égalité a lieu sur un certain voisinage de ∞.

Si A_n est de rang fini, le comportement global du spectre ne change pas grâce à (1.3), cependant le bord du spectre peut se déformer. Si θ est la plus grande valeur propre de A_n et $\mathbb{E}|X_{1,2}|^2 = 1$, en probabilité,

$$\lambda_{W_n} \underset{n \to +\infty}{\longrightarrow} \begin{cases} \theta + \frac{1}{\theta} & \text{si } \theta > 1, \\ 2 & \text{sinon.} \end{cases}$$

Ce phénomène de palier, qui montre que la plus grande valeur propre de la matrice déformée se détache du bord du spectre limite seulement si la perturbation dépasse un certain seuil porte le nom de *transition BBP*, du nom des premiers auteurs à avoir découvert ce phénomène, Ben Arous, Baik et Péché [11]. Les résultats les plus fins sur ce phénomène repose sur un calcul de déterminant développé par Benaych-Georges et Nadakuditi [20]. Il consiste à écrire $A = VDV^*$ où D est diagonale de taille r, contenant les valeurs propres non nulles et V de taille $n \times r$ est la matrice d'une famille orthonormée de vecteurs propres associés aux valeurs propres non nulles. La plus grande valeurs propre de W_n, si elle n'appartient pas au spectre de $X/\sqrt{n}$ est alors le plus grand zéro du déterminant,

$$\det(x - W_n) = \det G \det(I_n - GVDV^*),$$

où $G = (x - X/\sqrt{n})^{-1}$. En utilisant la formule de Frobenius $\det(I - AB) = \det(I - BA)$, on se ramène à un déterminant d'une matrice de taille r,

$$\det(x - W_n) = \det G \det(I_r - V^* GDV). \tag{1.7}$$

Informellement, puisque r est fixe, on peut utiliser l'asymptotique de la résolvante $G \simeq g_{\mu_{sc}}(x) I_n$, pour $x > \lambda_{X/\sqrt{n}}$, ce qui donne

$$\det(I_n - GVDV^*) \simeq \det(I_r - g_{\mu_{sc}}(x) D). \tag{1.8}$$

Puisque $g_{\mu_{sc}} : [2, +\infty) \to (0, 1]$ est décroissante, on en déduit que le plus grand zéro de $\det(x - W_n)$ est proche de $g_{\mu_{sc}}^{-1}(1/\theta)$, si $\theta > 1$, autrement dit de $\theta + 1/\theta$ par (1.5).

Une autre approche au comportement du bord du spectre sous des perturbations de rang fini est donnée par les probabilités libres. En effet, le bord du support de

$\mu_{sc} \boxplus \mu_{A_n}$, b_{s+A_n} s'exprime grâce à un résultat de Biane [23], $b_{s+A_n} = u_{A_n} + g_{A_n}(u_{A_n})$, où g_{A_n} désigne la transformée de Stieltjes de μ_{A_n} et,

$$u_{A_n} = \sup\{t \in \mathbb{R} : -g'_{A_n}(t) \geq 1\}.$$

On peut se convaincre si A_n est de rang fini que $u_{A_n} \simeq \theta$ si $\theta > 1$. Puisque $g_{A_n}(z) \simeq 1/z$, on en déduit

$$b_{s+A_n} \simeq \theta + \frac{1}{\theta}.$$

Cette approche a été développée et raffinée, en particulier pour comprendre le comportement des *outliers* des modèles déformés plus généraux, c'est-à-dire des valeurs propres se détachant du support de la mesure spectrale limite (voir [15] par exemple).

1.6 Fluctuations des statistiques linéaires

A partir du Théorème de Wigner, on peut se poser la question des fluctuations de la mesure spectrale empirique. Des arguments de concentration (voir [3, Théorème 2.3.5]) montrent que si X est une matrice de Wigner à entrées sous-Gaussiennes, alors pour toute fonction 1-Lipschitz et pour tout $t > 0$,

$$\mathbb{P}\big(\mathrm{tr} f(X/\sqrt{n}) - \mathbb{E} f(X/\sqrt{n}) > t\big) \leq e^{-ct^2},$$

où c est une certaine constante strictement positive. On en déduit la tension de la suite,

$$n(\mu_{X/\sqrt{n}}(f) - \mathbb{E}\mu_{X/\sqrt{n}}(f)).$$

Le fait que le facteur de “zoom” soit n et non pas $\sqrt{n}$ comme dans le cas des moyenne empirique de variables indépendantes, est une manifestation à l'échelle macroscopique de la répulsion entre les valeurs propres. Une première difficulté pour donner un sens à la convergence en loi, est de trouver un bon espace dans lequel plonger $n(\mu_{X/\sqrt{n}} - \mathbb{E}\mu_{X/\sqrt{n}})$. Généralement, on considère le processus $n(\mu_{X/\sqrt{n}}(f) - \mathbb{E}\mu_{X/\sqrt{n}}(f))_{f \in \mathcal{A}}$, où $\mathcal{A}$ est un sous-espace de $\mathcal{C}_b(\mathbb{R})$, l'ensemble des fonction continues bornées sur $\mathbb{R}$, que l'on souhaite le plus large possible.

Un des premiers résultats de théorème central limite remonte à Jonsson [62], repris dans [3, Théorème 2.1.31], et donne, en utilisant la méthode des moments, les fluctuations des traces de puissance de matrice de Wigner vers une loi Gaussienne de variance explicite dépendant seulement du $4^{\text{ème}}$ moment des entrées. Les fluctuations Gaussiennes des moments entraînent par des arguments de troncations, celles du processus indexés par les fonctions analytiques sur un domaine contenant le support de la loi du semi-cercle (voir par exemple [89]).

Une autre méthode classique, consiste à étudier le processus indexé par les fonctions du type $f_z(x) = (z - x)^{-1}$ pour $z \in \mathbb{C}^+$ et de montrer par la méthode des martingales, la convergence vers un processus Gaussien. La formule de Cauchy permet alors d'étendre cette convergence aux fonctions analytiques. Un résultat de Bai et Yao [9] présente un traitement complet du processus indexé par des fonctions analytiques, avec une fonction de moyenne et de covariance explicite.

D'autre part, on mentionne une autre technique utilisé dans [19], dans la cadre dans matrices de Wigner à queues lourdes, pour étendre la convergence en loi, à partir de celle du processus indexé par les fonctions f_z, $z \in \mathbb{C}^+$, à une classe plus large de fonctions test, grâce à la formule d'Helffer-Sjöstrand [3, (5.5.11)], qui permet de reconstruire une mesure à partir de sa transformée de Stieltjes. Si φ est une fonction C^k à support compact, et $\mu \in \mathcal{P}(\mathbb{R})$,

$$\int \varphi d\mu = \int_{\mathbb{C}^+} \Re\big(\overline{\partial}\Phi(z) g_\mu(z)\big) dz,$$

où Φ est une certaine extension de φ à $\mathbb{C}^+$ telle que $\overline{\partial}\Phi(z) = \frac{1}{k!}\varphi^{(k)}(\Re z)(\Im z)^k$, sur un voisinage de $\mathbb{R}$.

1.7 Ensembles Gaussiens classiques

Les ensembles Gaussiens classiques sont des modèles fondamentaux de matrices aléatoires en particulier parce qu'ils vérifient des propriétés de symétrie et forment des modèles intégrables. Pour cette raison, ils constituent des modèles de référence dans les approches aux questions d'universalité qui sont principalement basées sur des arguments de comparaison à ces ensembles.

On dit qu'une matrice $X \in \mathcal{H}_n^{(\beta)}$ appartient à l'*ensemble orthogonal Gaussien* (GOE) dans le cas où $\beta = 1$, et à l'*ensemble unitaire Gaussien* (GUE) pour $\beta = 2$, si elle suit la loi,

$$\frac{1}{Z_n^{(\beta)}} e^{-\frac{\beta}{4}\operatorname{tr} H^2} d\ell_n^{(\beta)}(H),$$

où $\ell_n^{(\beta)}$ désigne la mesure de Lebesgue sur $\mathcal{H}_n^{(\beta)}$, et $Z_n^{(\beta)}$ est la constante de normalisation. Le facteur $\beta/4$ est choisi de façon à ce que les entrées hors-diagonales soient de variance 1 (dans le cas complexe $\mathbb{E}|X_{1,2}|^2 = 1$).

1.7.1 Invariance sous le groupe $\mathcal{U}_n^\beta$

Notons $\mathcal{U}_n^\beta$ le groupe orthogonal pour $\beta = 1$ et unitaire pour $\beta = 2$. La loi de ses modèles de matrices est invariante sous l'action du groupe $\mathcal{U}_n^\beta$, et est caractérisée par cette invariance : *toute matrice de Wigner telle que* $\mathbb{E}|X_{1,2}|^2 = 1$*, invariante en loi sous l'action de* $\mathcal{U}_n^\beta$*, appartient au GOE si* $\beta = 1$ *ou au GUE si* $\beta = 2$.

De la même façon qu'une mesure de probabilité produit sur $\mathbb{R}^n$ invariante par rotation est nécessairement une loi Gaussienne (voir [63, Théorème 11.2]), on peut observer dans le cas $\beta = 1$, que si X est une matrice symétrique réelle invariante en loi sous l'action du groupe orthogonal, en testant l'invariance par conjugaison par

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{et} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \oplus I_{n-2},$$

on obtient

$$X_{1,1} \sim \frac{1}{2}\Big(X_{1,1}^{(1)} + X_{1,1}^{(2)} + X_{1,1}^{(3)} + X_{1,1}^{(4)}\Big), \quad X_{1,2} \sim \frac{X_{1,1} + X_{2,2}}{2},$$

où $X_{1,1}^{(i)}$ sont des copies indépendantes de $X_{1,1}$. La première égalité en loi donne une équation fonctionnelle pour la transformée de Fourier de $X_{1,1}$ qui oblige $X_{1,1}$ à suivre une loi Gaussienne centrée de variance σ^2. La deuxième égalité en loi permet de conclure que $X_{1,2} \sim \mathcal{N}(0,1)$ et $X_{1,1} \sim \mathcal{N}(0,2)$.

Plus généralement on définit des modèles orthogonalement/unitairement invariant de matrices aléatoires associées à un potentiel V, par la mesure de probabilité

$$\frac{1}{Z_V^n} e^{-n\mathrm{tr}V(H)} d\ell_n^{(\beta)}(H). \tag{1.9}$$

Pour donner sens à cette mesure de probabilité, on suppose que V satisfait la propriété de croissance suivante,

$$\liminf_{|x|\to+\infty} \frac{V(x)}{\beta' \log|x|} > 1, \tag{1.10}$$

pour un certain $\beta' \geq \beta, \beta' > 1$. L'invariance en loi sous $\mathcal{U}_n^\beta$ implique (voir [3, Corollaire 2.5.4]) que les vecteurs propres sont Haar-distribués sur $\mathcal{U}_n^\beta$, en particulier ils sont complètement délocalisés. L'invariance sous $\mathcal{U}_n^\beta$ permet aussi de calculer explicitement la loi du spectre de ces modèles de matrices aléatoires.

1.7.2 Loi du spectre des modèles unitairement invariant

La loi jointe des valeurs propres d'une matrice X distribuée selon (1.9) est donnée par,

$$\mathbb{P}_{V,\beta}^n = \frac{1}{Z_{V,\beta}^n} |\Delta(\lambda)|^\beta e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i, \tag{1.11}$$

où $Z_{V,\beta}^n$ est la fonction de partition, et Δ désigne le déterminant de Vandermonde,

$$\Delta(\lambda) = \prod_{i<j} (\lambda_i - \lambda_j).$$

Ce déterminant de Vandermonde provient du Jacobien de l'application

$$(U, D) \in \mathcal{U}_n^\beta \times \mathbb{R}^n \mapsto U^* D U,$$

et révèle le comportement de répulsion logarithmique entre les valeurs propres des modèles orthogonalement ou unitairement invariants.

1.7.3 Représentation tridiagonale

Pour $\beta > 0$, $\mathbb{P}_{\beta,V}^n$ définit un gaz de Coulomb à la température $1/\beta$ associé au potentiel V, et à l'interaction 2-dimensionnelle logarithmique restreinte à $\mathbb{R}$. Lorsque V est quadratique $V = \beta x^2/4$, $P_{\beta,V}^n$ peut s'interpréter, grâce à un résultat de Dimitru-Edelman [46], [3, Théorème 4.5.35], comme la loi jointe des valeurs propres du modèle tridiagonal, $J_n = X/\sqrt{\beta n}$, où

$$X = \begin{pmatrix} \xi_1 & \zeta_1 & 0 & \cdots & 0 \\ \zeta_1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \zeta_{n-1} \\ 0 & \cdots & 0 & \zeta_{n-1} & \xi_n \end{pmatrix},$$

et $(\xi_i, \zeta_i)_{i\geq 1}$ est une famille de variables indépendantes, telle que $(\xi_i)_{i\geq 1}$ sont des variables Gaussiennes de variance 2 et ζ_i suit une χ-distribution de degré $i\beta$. Cette représentation établit un lien entre les matrices aléatoires et les opérateurs de Schrödinger aléatoires sur $\mathbb{Z}$. L'avantage de cette approche est qu'elle permet d'utiliser une *méthode objective* dans l'étude du comportement du spectre des ensembles Gaussiens classiques, dans le sens où Aldous et Steele l'entendent dans [2]. La représentation comme opérateur de Schrödinger du GOE et GUE permet de donner un sens à la limite d'opérateur quand n tend vers $+\infty$ vers un opérateur infini-dimensionnel dont on peut déduire certaines propriétés asymptotiques du spectre du GOE ou GUE.

Par exemple, on peut utiliser cette approche pour redémontrer le théorème de Wigner, comme le suggère Virág [100]. On note $(X_k, Y_k)_{1\leq k\leq n}$ les coefficients diagonaux et hors-diagonaux de J_n (avec la convention $Y_n = 0$). On voit J_n comme une mesure de probabilité stationnaire sur $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$, notée ρ_n, en considérant la loi de $(X_{U+k[n]}, Y_{U+k[n]})_{k\in\mathbb{Z}}$ où U est uniformément échantillonnée dans $\{1, ..., n\}$. Ceci revient à voir J_n comme la loi de la matrice de Jacobi infinie à coefficients stationnaires $(X_{U+k[n]}, Y_{U+k[n]})_k$. Notons $\mathcal{P}_{\text{stat}}$ l'ensemble des mesures de probabilité stationnaires sur $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$, muni de sa topologie produit. On peut associer à un élément $\rho \in \mathcal{P}_{\text{stat}}$, vu comme la loi d'une matrice de Jacobi à coefficients stationnaires, une mesure spectrale naturelle μ_ρ par la formule,

$$\mu_\rho = \mathbb{E}_\rho \mu_J^{e_o},$$

où sous $\mathbb{P}_\rho$, J suit la loi ρ, et où $\mu_J^{e_o}$ désigne la mesure spectrale associée au vecteur e_o. En particulier, on voit que, $\mu_{\rho_n} = \mu_n$, la mesure spectrale empirique de J_n. Ces définitions sont des versions simplifiées des notions de topologie locale faible et de mesure spectrale associée aux mesures unimodulaires dans le cadre de graphes sparse (voir [2] et [28] pour une introduction). L'avantage de cette représentation est que l'application $\rho \mapsto \mu_\rho$ est continue pour la topologie faible par [31, Théorème 2.2]. Revenant à la représentation tridiagonale du GOE/GUE, on peut se convaincre, puisque $\mathbb{E}\zeta_k = \sqrt{k - 1/\beta}$,

$$\mathbb{E}\rho_n \underset{n\to+\infty}{\rightsquigarrow} \rho_\infty,$$

où ρ_∞ est la loi de la matrice de Jacobi infinie de coefficients diagonaux nuls et hors-diagonaux tous égaux à $\sqrt{U}$ où U est uniforme sur $[0, 1]$. D'autre part,

$$\mu_{\mathbb{E}\rho_n} = \mathbb{E}\mu_n,$$

et $\mu_{\rho_\infty} = \mu_{sc}$. On en déduit par continuité de la mesure spectrale,

$$\mathbb{E}\mu_n \underset{n\to+\infty}{\rightsquigarrow} \mu_{sc}.$$

1.7.4 Structure déterminantale

Dans le cas $\beta = 2$, les modèles unitairement invariant possèdent la caractéristique fondamentale que le processus de leurs valeurs propres a une structure déterminantale. En effet, la densité p_V^n du vecteur de ses valeurs propres, peut s'écrire sous la forme,

$$p_V^n(\lambda_1, ..., \lambda_n) = \frac{1}{n!}\det(K_n(\lambda_i, \lambda_j))_{1\leq i,j\leq n},$$

où K_n est le noyau,

$$K_n(x, y) = \sum_{i=1}^{n} \psi_i(x)\psi_j(y)e^{-\frac{V(x)}{2}-\frac{V(y)}{2}},$$

où ψ_i désignent les polynômes orthogonaux associés à $e^{-V(x)}dx$. Comme le noyau est auto-reproduisant, c'est-à-dire,

$$\forall x, y \in \mathbb{R}, \int K_n(x, t)K_n(t, y)dt = K_n(x, y),$$

le processus des valeurs propres est bien déterminantal (voir [60, Exercice 4.1.1]). L'avantage de cette formulation est qu'elle permet d'exprimer les probabilités de trou du processus comme des déterminants de Fredholm associé au noyau K_n (voir [3, Définition 3.4.3]) qui satisfont de bonnes propriétés de continuité par rapport au noyau (voir [3, Lemme 3.4.5]). Les différentes propriétés asymptotiques du processus ponctuel des valeurs propres peuvent alors se déduire de asymptotique du noyau qui gouverne le processus déterminantal, c'est-à-dire des polynômes orthogonaux ψ_n. Revenant au cas du GUE, on distingue deux cas de figure : la convergence du processus vu du *bulk*, et celui vu du bord du spectre.

Le théorème de Wigner nous dit que l'espacement moyen au milieu du spectre (la densité du semi-cercle étant localement constante) entre les valeurs propres d'une matrice de Wigner est de l'ordre de $1/\sqrt{n}$, ce qui indique que le processus pertinent à regarder en ce concerne le comportement dans le bulk est

$$\Xi_{\text{bulk}} = \sum_{i=1} \delta_{\sqrt{n}\lambda_i},$$

où $\lambda_1 \leq ... \leq \lambda_n$ sont les valeurs propres du GUE. On considère Ξ_{bulk} comme une variable aléatoire dans l'espace des mesures de Radon atomiques, muni de la topologie engendrée par les fonctions continues à support compact. L'asymptotique des polynômes de Hermite permet d'obtenir le théorème suivant dû à Metha et Gaudin (voir [3, Théorème 3.1.1], [41, §5.5]),

1.7.1 Théorème. *On définit le noyau sinus,*

$$\forall x, y \in \mathbb{R}, \; K_{sine}(x, y) = \frac{\sin(x-y)}{\pi(x-y)},$$

On note Ξ_{sine} le processus déterminantal associé à ce noyau. Le processus Ξ_{bulk} converge en loi vers le processus sinus Ξ_{sinus}.

L'existence du processus sinus est assuré par le fait que l'opérateur intégral associé au noyau K_{sine} sur $L^2_{\text{loc}}(\mathbb{R})$ est un opérateur projection (dans le domaine de Fourier il correspond à l'opérateur de multiplication par $\mathbb{1}_{[-\pi,\pi]}$, voir [60, Théorème 4.5.3 et §4.3.5]).

En ce qui concerne le bord du spectre, on peut considérer informellement que les valeurs propres du GUE normalisée par $\sqrt{n}$ doivent être proche des quantiles de la loi du semi-cercle, mais l'avant-dernier $1/n$-quantile γ_{n-1}

$$\int_{\gamma_{n-1}}^{2} \frac{1}{2\pi}\sqrt{4-x^2}dx = \frac{1}{n},$$

est tel que $2-\gamma_{n-1} \asymp n^{-2/3}$. De façon à voir un nombre d'ordre 1 de valeurs propres dans un compact, on considère donc le processus

$$\Xi_{\text{edge}} = \sum_{i=1}^{n} \delta_{\widetilde{\lambda}_i}, \quad \widetilde{\lambda}_i = n^{3/2}\Big(\frac{\lambda_i}{\sqrt{n}} - 2\Big).$$

Avec cette normalisation, on a la convergence suivante du processus (voir [3, Théorème 3.1.4]),

1.7.2 Théorème. *Soit*

$$\forall x,y \in \mathbb{R},\ K_{Airy}(x,y) = \frac{Ai(x)Ai'(y) - Ai(y)Ai'(x)}{x-y},$$

où Ai désigne la fonction d'Airy, unique solution de d'équation différentielle,

$$\frac{d^2y}{dx^2} - xy = 0,$$

avec comme conditions initiales,

$$Ai(0) = \frac{1}{3^{2/3}\Gamma(2/3)}, \quad Ai'(0) = \frac{1}{3^{1/3}\Gamma(1/3)}.$$

On note Ξ_{Airy} le processus déterminantal de noyau K_{Airy}. Le processus Ξ_{edge} converge en loi vers Ξ_{Airy}.

En particulier, on obtient la convergence de la fonction de répartition de $n^{2/3}(\lambda_n - 2)$ vers la fonction F_2,

$$F_2(t) = \mathbb{P}\big(\Xi_{\text{Airy}}(t,+\infty) = 0\big),$$

qui a été identifiée par Tracy et Widom [97] comme étant la fonction de répartition d'une certaine loi, appelée loi de Tracy-Widom TW_2. Informellement, TW_2 est la loi de la plus grande particule du processus d'Airy.

Dans le cas où $\beta > 0$, la formulation déterminantale n'est plus valable, cependant la représentation tridiagonale a permis à Ramirez, Rider et Virag [83] de donner un sens à la limite d'opérateurs de

$$n^{2/3}\big(J_n - 2I_n\big),$$

vers un opérateur limite appelé opérateur stochastique d'Airy. Par la méthode objective, ils en déduisent la convergence de $n^{2/3}(\lambda_n - 2)$ vers la loi de Tracy-Widom TW_β et identifient les exposants des queues de distribution de TW_β,

$$TW_\beta(-\infty, -t] = e^{-\frac{1}{24}\beta a^3(1+o(1))},$$

$$TW_\beta[t, +\infty)- = e^{-\frac{2}{3}\beta a^{3/2}(1+o(1))}.$$

L'asymétrie des queues de distribution de TW_2 vient du fait que les déviations à gauche de la plus grande valeur propre porte en elle les déviations de tout le spectre.

Cette loi apparaît dans de nombreux problèmes de combinatoire dont les deux plus célèbres sont les problèmes de la plus longue sous-suite croissante et de percolation de dernier passage.

Ulam's problem. Baik, Deift et Johansson [12], ont montré que la plus longue sous-suite croissante L_n d'une permutation uniformément échantillonnée dans $\mathfrak{S}_n$, l'ensemble des permutations de $\{1, ..., n\}$, a le même comportement que celui de la plus grande valeur propre du GUE, c'est-à-dire,

$$n^{2/3}\Big(\frac{L_n}{\sqrt{n}} - 2\Big) \underset{n\to+\infty}{\rightsquigarrow} TW_2.$$

D'autre part, le *temps de percolation de dernier passage* $T(X)$,

$$T(X) = \sup_{\pi} \sum_{v\in\pi} X_v,$$

où le supremum porte sur l'ensemble des chemins de $(1,1)$ à (n,n), de pas $(1,0)$ ou $(0,1)$, pour une famille de poids i.i.d $(X_{i,j})_{1\leq i,j\leq n}$ de loi exponentielle, donne un modèle intégrable, pour lequel on sait par Johansson [61],

$$n^{2/3}\Big(\frac{T(X)}{n} - 4\Big) \underset{n\to+\infty}{\rightsquigarrow} TW_2.$$

Lorsque les poids satisfont une condition d'intégrabilité légèrement plus forte que l'existence d'un moment d'ordre 2, il est connu que $T(X)/n$ converge vers une limite déterministe γ (voir [75]), que l'on ne sait identifier que dans le cas exponentiel. Il est conjecturé, lorsque les poids sont suffisamment intégrables, que les fluctuations sont universelles et qu'elles sont données par la loi de Tracy-Widom.

1.8 Grandes déviations

Dans toute la suite on s'intéressera à des problèmes de grandes déviations qui apparaissent dans le contexte des matrices de Wigner et des β-ensembles.

1.8.1 Principe de grandes déviations

On dit qu'une suite de mesures de probabilité μ_n sur un espace topologique $\mathcal{X}$ muni de sa tribu Borélienne $\mathcal{B}$, suit un *principe de grandes déviations* (PGD), à vitesse $v : \mathbb{N} \to \mathbb{N}$, et fonction de taux $J : \mathcal{X} \to [0, +\infty]$, si J est semi-continue inférieurement, v tend vers $+\infty$ et pour tout $B \in \mathcal{B}$,

$$-\inf_{B^\circ} J \leq \liminf_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(B) \leq \limsup_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(B) \leq -\inf_{\overline{B}} J,$$

où B° désigne l'intérieur de B et $\overline{B}$ la fermeture de B. Si de plus, tout les sous-niveaux de J sont compacts on dira que J est une *bonne fonction de taux.* Souvent, on ne fera pas de distinction dans les énoncés de PGD entre une suite de variables aléatoires et la suite de leurs lois.

Si $\mathcal{X}$ est un espace métrique, l'hypothèse de semi-continuité inférieure de la fonction de taux permet de dire, puisque pour tout $x \in \mathcal{X}$,

$$\sup_{\delta>0} \inf_{B(x,\delta)} J = J(x),$$

que l'on a,

$$\lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(B(x,\delta)) = -J(x),$$

et de même pour la $\liminf$. Ceci permet d'interpréter J comme étant une sorte de densité à l'échelle exponentielle par rapport à une mesure de référence $\mu_n \simeq e^{-v(n)J}\mu$. Le *lemme de Varadhan* permet en quelque sorte de donner une réciproque à cette interprétation. Si $(\nu_n)_{n\in\mathbb{N}}$ est une suite de mesures de probabilité satisfaisant un PGD à vitesse $v(n)$ et de fonction de taux J, et $\varphi : \mathcal{X} \to \mathbb{R}$ est une fonction continue bornée, alors la suite

$$\mu_n = \frac{1}{Z_n} e^{v(n)\varphi} \nu_n,$$

où Z_n est une constante de normalisation, satisfait un PGD à vitesse v et de fonction de taux,

$$J - \varphi - \inf\{J - \varphi\}.$$

Par ailleurs, la semi-continuité inférieure de la fonction de taux implique que les valeurs d'adhérence de μ_n ont leur support inclus dans $\{J = 0\}$.

Généralement, la stratégie adoptée pour montrer un PGD consiste à commencer par prouver un *PGD faible*, c'est-à-dire, montrer la borne inférieure pour des ouverts et la borne supérieure pour des ensembles compacts. Lorsque $\mathcal{X}$ est un espace métrique, on montre un PGD faible en estimant à l'échelle exponentielle la probabilité de petites boules $B(x, \delta)$,

$$\liminf_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(B(x,\delta)) \geq -J(x),$$

$$\lim_{\delta\to 0} \limsup_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(B(x,\delta)) \leq -J(x).$$

Puis on montre que la suite μ_n est *exponentiellement tendue*, c'est-à-dire qu'il existe une suite de compacts K_m telle que

$$\lim_{m\to+\infty} \limsup_{n\to+\infty} \frac{1}{v(n)} \log \mu_n(K_m^c) = -\infty,$$

pour en conclure un PGD complet.

La notion de bonne fonction de taux a l'avantage qu'elle permet de dire qu'un infimum sur un fermé est nécessairement atteint. En particulier si F est un fermé inclus dans $\{J > 0\}$, alors

$$\mu_n(F) \underset{n\to+\infty}{\longrightarrow} 0.$$

1.8.2 Grandes déviations des β-ensembles

Le premier résultat de grandes déviations dans le cadre des matrices aléatoires a été obtenu par Ben Arous et Guionnet [4, Théorème 1.3] pour la mesure spectrale empirique des β-ensembles associés à une potentiel V continu et satisfaisant la condition de croissance (1.10).

1.8.1 Théorème. *La mesure empirique*

$$L_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i},$$

suit sous $\mathbb{P}^n_{\beta,V}$ un PGD à vitesse n^2 et bonne fonction de taux I_β, definie pour tout $\mu \in \mathcal{P}(\mathbb{R})$ par,

$$I_\beta(\mu) = \int V(x)d\mu(x) - \beta\Sigma(\mu) - c_\beta,$$

où $\Sigma(\mu)$ désigne l'entropie non-commutative de μ et est définie par,

$$\Sigma(\mu) = \begin{cases} \int \log|x-y|d\mu(x)d\mu(y) & \textit{si } \int \log(1+|x|)d\mu(x) < +\infty, \\ -\infty & \textit{sinon,} \end{cases}$$

et où c_β est une constante telle que $\inf I_\beta = 0$.

La fonction de taux I_β s'annule en une unique mesure de probabilité σ_β^V, appelée mesure d'équilibre. A cause de l'hypothèse (1.10), le caractère confinant du potentiel l'emporte sur la répulsion des particules, ce qui implique que la mesure d'équilibre σ_β^V est à support compact. Comme $\Sigma(\sigma_\beta^V) < +\infty$, on en déduit par [88, Proposition A 6], que σ_β^V ne met aucune masse sur des ensembles de capacité nulle. De plus, σ_β^V est caractérisée par

$$V(x) - p_{\sigma_\beta^V}(x) = C_\beta^V, \tag{1.12}$$

pour σ_β^V-presque tout x et par le fait que

$$\forall x \notin \mathrm{supp}\sigma_\beta^V,\ V(x) - p_{\sigma_\beta^V}(x) > C_\beta^V,$$

pour une certaine constante C_β^V, et où $p_{\sigma_\beta^V}$ désigne le potentiel logarithmique de σ_β^V, défini par

$$p_{\sigma_\beta^V}(x) = \int \log|x-y|d\sigma_\beta^V(y).$$

Il est connu que le potentiel logarithmique d'une mesure à support compact prend la valeur $-\infty$ au plus sur un ensemble de capacité nulle (voir [88, Corollaire A 5]), ce qui donne bien un sens aux deux conditions énoncées plus haut.

Puisque σ_β^V est l'unique minimiseur de I_β qui est une bonne fonction de taux, on en déduit,

$$L_n \underset{n\to+\infty}{\rightsquigarrow} \sigma_\beta^V,$$

en probabilité.

On peut relaxer l'hypothèse sur la croissance du potentiel, et autoriser, grâce au résultat de Hardy [59], des potentiels qui ont une croissance comparable au logarithme. La mesure d'équilibre obtenue alors n'est plus nécessairement à support compact.

L'idée de la preuve du Théorème 1.8.1 est de commencer par observer que $\mathbb{P}^n_{V,\beta}$ peut se réécrire sous la forme,

$$\mathbb{P}^n_{\beta,V} = \frac{1}{Z^n_{V,\beta}} e^{-n^2 \int_{x\neq y} f(x,y)dL_n(x)dL_n(y)} \mu_V^n,$$

où

$$f(x,y) = \frac{1}{2}V(x) + \frac{1}{2}V(y) - \beta \log|x-y|,$$

et $\mu_V = Z_V^{-1} e^{-V(x)}dx$. Puisque sous μ_V^n, la mesure empirique L_n a des déviations à vitesse n par le Théorème de Sanov (voir [43, Theorem 6.2.10]), on pourrait en déduire si l'application,

$$\mu \mapsto \int_{x\neq y} f(x,y)d\mu(x)d\mu(y), \tag{1.13}$$

était continue pour la topologie faible, par le lemme de Varadhan, le résultat du Théorème 6.1.1, avec pour fonction de taux la fonction (1.13) (translatée d'une certaine constante). Cependant on peut voir facilement que l'application (1.13) n'est pas continue, en prenant une suite de mesures atomiques qui converge vers δ_0, dont les atomes s'approchent suffisamment vite pour faire diverger le logarithme.

La preuve du Théorème 6.1.1 consiste alors à utiliser la méthode de Laplace, en faisant attention dans la borne inférieure à contourner la singularité du logarithme.

L'entropie non-commutative qui apparaît dans ce PGD permet de créer certains liens avec la notion de capacité. En particulier, on sait d'après [88, Théorème A 20], que tout sous-ensemble compact de $\mathbb{R}$ de dimension de Hausdorff strictement positive est de capacité strictement positive, ce qui veut dire que pour tout compact de dimension de Hausdorff strictement positive, il existe une mesure supportée sur ce compact qui est d'entropie non-commutative finie. On en déduit que le domaine de la fonction de taux du Théorème 6.1.1 contient des mesures très singulières.

De nombreux résultats plus fins prolongent ce résultat de grandes déviations. En particulier, on mentionne les développements asymptotiques de la fonction de partition, progressivement étendus à tout ordre, obtenus par Borot et Guionnet [34]. Un PGD pour le processus ponctuel du bulk étiqueté et moyenné, c'est-à-dire le processus vu d'un point du support Σ de σ_β^V, enrichi de l'information du point observateur, puis moyenné,

$$\overline{\Xi}_{\text{bulk}} = \frac{1}{|\Sigma|}\int_\Sigma \sum_{i=1}^n \delta_{x,n(x-\lambda_i)}dx,$$

où $|\Sigma|$ désigne la mesure de Lebesgue de Σ, a été obtenu par Leblé et Serfaty [68]. En particulier, leur résultat permet de voir le processus sinus comme solution d'un principe variationnel.

En ce qui concerne les grandes déviations de la plus grande particule sous $\mathbb{P}^n_{V,\beta}$, on a le résultat suivant dû à Ben Arous, Dembo et Guionnet [16].

1.8.2 Théorème. *On suppose que*

$$\lim_{n\to+\infty} \frac{1}{n} \log \frac{Z^{n-1}_{\frac{nV}{n-1},\beta}}{Z^n_{V,\beta}} = C^V_\beta,$$

où C^V_β est définie par (1.12). La plus grande particule $\max_i \lambda_i$ suit sous $\mathbb{P}^n_{V,\beta}$, un PGD à vitesse n et de bonne fonction de taux J_β définie pour $x \geq \lambda^* = \sup \operatorname{supp} \sigma^V_\beta$,

$$J_\beta(x) = V(x) - \beta \int \log|x-y| d\sigma^V_\beta(y) - C^V_\beta,$$

et pour $x < \lambda^*$, $J_\beta(x) = +\infty$.

L'hypothèse sur la convergence du ratio des fonctions de partition est en fait équivalente à la convergence

$$\max_i \lambda_i \underset{n\to+\infty}{\longrightarrow} \lambda^*,$$

en probabilité sous $\mathbb{P}^n_{V,\beta}$. La preuve du Théorème 1.8.2 repose sur le fait qu'en faisant un changement de mesure, on peut considérer que les $n-1$ plus petites particules forment un β-ensemble associé au potentiel $\frac{n}{n-1}V$, dont les déviations sont négligeables à l'échelle n par le Théorème 6.1.1. Le changement de mesure fait intervenir le ratio des fonctions de partition, et donne à l'échelle exponentielle la constante de normalisation de la fonction de taux. Le terme d'interaction entre la plus grande valeur propre et les $n-1$ plus petites fournit la fonctionnelle de taux (à une constante près). On note que l'hypothèse de convergence du ratio des fonction de partition vers C^V_β est cruciale pour pouvoir définir une candidate pour la fonction de taux qui soit positive.

1.8.3 Grandes déviations des modèles déformés

A partir des différents résultats de convergence du spectre des modèles déformés, on peut se poser la question des grandes déviations de leur mesure spectrale empirique et de leur plus grande valeur propre. Si on considère une matrice X du GUE ($\beta = 2$) ou GOE ($\beta = 1$), et une suite de matrices déterministes A_n, la matrice déformée

$$W_n = X/\sqrt{n} + A_n,$$

admet une expression explicite pour la loi jointe de ses valeurs propres, qui fait intervenir des intégrales sphériques $I^{(\beta)}(D_n, A_n)$,

$$|\Delta(\lambda)|^\beta I^{(\beta)}(D_n, A_n) e^{-\frac{\beta}{4} n \mathrm{tr} A_n^2} e^{-\frac{\beta}{4} n \sum_{i=1}^n \lambda_i^2} \prod_{i=1}^n d\lambda_i, \tag{1.14}$$

où D_n est la matrice diagonale formées de $(\lambda_1, ..., \lambda_n)$ et

$$I^{(\beta)}(D_n, A_n) = \int_{\mathcal{U}_n^\beta} e^{\frac{\beta}{2}\mathrm{tr} U D_n U^* A_n} dU,$$

où dU désigne la mesure de Haar sur $\mathcal{U}_n^\beta$. Dans le régime où A_n est de rayon spectral uniformément borné et où sa mesure spectrale empirique converge vers une mesure à support compact, la stratégie adoptée par Guionnet et Zeitouni [58] pour comprendre les grandes déviations de la mesure spectrale empirique de W_n repose sur une approche dynamique qui consiste à considérer le processus de la mesure spectrale empirique associée au mouvement Browninen de Dyson démarré en A_n,

$$Z_t = H_t/\sqrt{n} + A_n,$$

et à tirer avantage des outils de calcul stochastique pour obtenir un PGD pour le processus de la mesure spectrale de Z_t. Puisque $Z_1 = W_n$, ils en déduisent par contraction un PGD pour μ_{W_n} à vitesse n^2 et de fonction de taux explicite. La caractéristique frappante de la preuve de ce PGD est qu'elle ne repose pas sur la connaissance de la loi jointe du spectre. Cependant elle utilise de façon cruciale la nature Gaussienne des entrées et l'invariance par le groupe unitaire. En contrepartie, au regard de la loi explicite du spectre de W_n (1.14), le PGD obtenu permet de donner l'asymptotique à l'échelle exponentielle n^2 des intégrales sphériques $I^{(\beta)}(D_n, A_n)$, pour deux suites déterministes D_n et A_n de mesure spectrale convergeant vers des mesures à support compact.

En ce qui concerne le bord du spectre, on peut se demander dans le régime où $\lim \mu_{A_n} = \delta_0$, quelles sont les grandes déviations de la plus grande valeur propre. Dans le cas où A_n est de rang 1, $A_n = \theta u u^*$, Maïda [78] obtient l'asymptotique des intégrales sphérique à l'échelle n, $I^\beta(D_n, \theta u u^*)$, uniformément en D_n de rayon spectral borné, ce qui permet d'analyser les grandes déviations à vitesse n de la plus grande valeur propre, et de donner une fonction de taux explicite qui illustre la transition BBP. Plus généralement, les grandes déviations des valeurs propres extrêmes de modèles unitairement invariant perturbés par des matrices de rang fini ayant des vecteurs propres complètement délocalisés ont été étudiées par Benaych-Georges, Guionnet et Maïda [18, Théorème 2.13]. Leur analyse repose sur la représentation des valeurs propres extrêmes des perturbations de rang fini comme zéros d'un certain déterminant d'une matrice de rang fini de la forme (1.8).

1.8.4 Matrices à entrées bornées

En dehors des modèles intégrables que l'on a mentionnés plus haut, les grandes déviations du spectre de matrices aléatoires restent en grande partie mal comprises. En particulier le cas des déviations des matrices à entrées bornées, de nature plus combinatoire, reste très largement ouvert.

Cependant, ces modèles de matrices présentent l'avantage d'être reliés naturellement à des modèles de graphes aléatoires, pour lesquels il est possible dans certains régimes de donner des PGD. On distingue généralement deux régimes : le régime dense où le nombre d'arêtes est proportionnel au carré du nombre de sommets, et le régime sparse où le degré de chaque sommet reste borné. Les deux régimes ont une théorie limite bien développée, mais seul le régime sparse dispose d'une théorie spectrale intéressante du point de vue des matrices aléatoires.

Dans le régime dense, Lovász et Szegedy ont développé la notion de *graphon* permettant de plonger tous les graphes denses dans un même espace (voir par exemple [74]). Plus précisément, un graphon est une classe d'équivalence d'applications mesurables symétriques $f : [0,1]^2 \to [0,1]$ sous l'action du groupe des bijections préservant la mesure de Lebesgue $\sigma : [0,1] \to [0,1]$, l'action étant définie par $\sigma.f(x,y) = f(\sigma(x),\sigma(y))$. Pour tout graphe fini (V,E), on peut lui associer le graphon naturel,

$$f_G(x,y) = \begin{cases} 1 & \text{si } (\lceil nx \rceil, \lceil nx \rceil) \text{ est une arrête,} \\ 0 & \text{sinon.} \end{cases}$$

Le segment $[0,1]$ peut se penser comme étant une compactification des sommets d'un graphe infini, qui se comprend alors comme un continuum de sommets. Le fait que l'on considère la classe d'équivalence sous l'action des bijections qui préserve la mesure correspond en quelque sorte au réétiquetage des sommets. On peut définir sur l'espace $\widetilde{W}$ des graphons une topologie qui correspond, si G_n est une suite de graphes finis convergeant vers $f \in \widetilde{W}$, à la convergence de toutes les "densités" de sous-graphes donnés vers celle du graphon f. D'un point de vu spectral, un graphon donne naissance naturellement à un opérateur de Hilbert-Schmidt sur $L^2([0,1])$. On peut montrer que l'application qui à un graphon associe le spectre de son opérateur d'Hilbert-Schmidt est continue pour la topologie des graphons (voir [39, Lemme 2.3] pour un énoncé plus précis).

Cependant, si on représente une matrice aléatoire à coefficients bornées X comme graphon puis comme opérateur d'Hilbert-Schmidt $\mathcal{K}$ sur $L^2([0,1])$, on peut voir que

$$\sigma(\mathcal{K}) = \sigma(X/n).$$

Dans ce régime, Chatterjee et Varadhan [39], [38] ont montré un PGD à vitesse n^2 pour le spectre de X/n, vu comme processus ponctuel où la topologie choisie est engendrée par les fonctions à support compact dans $\mathbb{R} \setminus \{0\}$. La preuve repose sur la contraction d'un PGD qu'ils obtiennent pour le graphon associé à X.

Dans le régime sparse, Aldous et Steele [2] ont introduit la notion de topologie locale faible, que l'on va essayer de présenter brièvement ainsi que son intérêt du point de vue des grandes déviations des matrices aléatoires. On considère $\mathcal{G}^*$ l'ensemble des classes d'équivalence de graphes localement finis enracinés, que l'on munit de la topologie locale engendrée par les fonctions $(g,o) \mapsto \mathbb{1}_{(g,o)_r \simeq (h,o)_r}$ où $(h,o) \in \mathcal{G}^*$ et $(g,o)_r$ désigne le r-voisinage de la racine dans g. La topologie locale encode la géométrie du graphe que voit la racine dans un horizon fini. On munit $\mathcal{P}(\mathcal{G}^*)$ de la topologie faible associée à la topologie locale sur $\mathcal{G}^*$, appelée topologie locale faible. Un graphe fini $G = (V,E)$ induit naturellement un élément de $\mathcal{P}(\mathcal{G}^*)$, en considérant

$$\rho_G = \frac{1}{|V|} \sum_{v \in V} \delta_{(G,v)},$$

où (G, v) désigne la classe d'équivalence du graphe G enraciné en v. Pour des mesures de probabilité ρ sur $\mathcal{G}^*$ dites unimodulaires, c'est-à-dire satisfaisant une certaine propriété de transport de masse, on peut définir une mesure spectrale qui coïncide avec la mesure spectrale empirique des valeurs propres lorsque $\rho = \rho_G$. Un des points fort de cette topologie est qu'elle est suffisamment fine pour que l'application $\rho \mapsto \mu_\rho$ soit continue sur l'ensemble des mesures unimodulaires, et en même temps suffisamment grossière pour pouvoir faire des grandes déviations.

En particulier, une stratégie de preuve peut consister à contracter un PGD d'une suite de graphes aléatoires G_n par rapport à la topologie locale faible, pour en déduire une PGD pour la mesure spectrale empirique de leurs matrices d'adjacence. Par exemple, grâce à un résultat de Bordenave et Caputo [30], on sait que le graphe d'Erdös-Renyi de paramètre c/n, $G(n, c/n)$, où $c > 0$, satisfait un PGD par rapport à la topologie locale faible, ce qui donne par contraction un PGD pour la mesure spectrale empirique de la matrice aléatoire symétrique X où $(X_{i,j})_{i \leq j}$ sont i.i.d de loi de Bernoulli de paramètre c/n.

1.8.5 Matrices de Wigner sans queues sous-Gaussiennes

Dans tous les PGD pour les matrices de Wigner que l'on a présenté précédemment, la loi des entrées était sous-Gaussienne. Lorsqu'on affaiblit l'intégrabilité des entrées, des phénomènes de queues lourdes peuvent apparaître, comme c'est le cas pour le modèle de *matrices de Wigner sans queues sous-Gaussiennes*.

1.8.1 Définition. On dit qu'une matrice de Wigner X est *sans queues sous-Gaussiennes* s'il existe $\alpha \in (0, 2)$ et $a, b \in (0, +\infty)$ tels que,

$$\lim_{t \to +\infty} -t^{-\alpha} \log \mathbb{P}\left(|X_{1,1}| > t\right) = b, \tag{1.15}$$

$$\lim_{t \to +\infty} -t^{-\alpha} \log \mathbb{P}\left(|X_{1,2}| > t\right) = a,$$

et deux mesures de probabilités sur $\mathbb{S}^1$, υ_1 et υ_2, et $t_0 > 0$, tels que pour tout $t \geq t_0$ et tout sous-ensemble mesurable U de $\mathbb{S}^1$,

$$\mathbb{P}\left(X_{1,1}/|X_{1,1}| \in U, |X_{1,1}| \geq t\right) = \upsilon_1(U)\mathbb{P}\left(|X_{1,1}| \geq t\right),$$

$$\mathbb{P}\left(X_{1,2}/|X_{1,2}| \in U, |X_{1,2}| \geq t\right) = \upsilon_2(U)\mathbb{P}\left(|X_{1,2}| \geq t\right). \tag{1.16}$$

1.8.3 Remarque. L'hypothèse sur la queue de distribution peut sembler assez restrictive, mais comme les déviations sont créées par des phénomènes de queues lourdes, c'est-à-dire par des grandes entrées de la matrice, la fonction de taux dépendra fortement de la queue de distribution exacte des entrées. La formulation de l'hypothèse (1.16) d'"indépendance" de l'angle et du module des entrées est faite pour s'assurer que l'on a des bornes inférieures de déviations des entrées, c'est-à-dire,

$$\lim_{\delta \to 0} \liminf_{n \to +\infty} n^{-\frac{\alpha}{2}} \log \mathbb{P}\big(X_{1,1}/\sqrt{n} \in B(x, \delta)\big) \geq -b|x|^\alpha,$$

pour $\mathrm{sg}(x) \in \mathrm{supp}(\nu_1)$, et

$$\lim_{\delta\to 0}\liminf_{n\to+\infty} n^{-\frac{\alpha}{2}} \log \mathbb{P}\big(X_{1,2}/\sqrt{n} \in B(z,\delta)\big) \geq -a|z|^{\alpha},$$

pour $z/|z| \in \mathrm{supp}(\nu_2)$.

Les grandes déviations de la mesure spectrale empirique de matrices de Wigner sans queues Gaussiennes ont été étudiées par Bordenave et Caputo [29]. Ils ont obtenu le résultat suivant.

1.8.4 Théorème. *Soit X une matrice de Wigner sans queues sous-Gaussiennes. La mesure spectrale empirique de $X/\sqrt{n}$ suit un PGD par rapport à la topologie faible. Le PGD est à la vitesse $n^{1+\alpha/2}$, et de bonne fonction de taux I_α définie pour tout $\mu \in \mathcal{P}(\mathbb{R})$, par*

$$I_\alpha(\mu) = \begin{cases} \Phi(\nu) & \text{si } \mu = \mu_{sc} \boxplus \nu \text{ avec } \nu \in \mathcal{P}(\mathbb{R}), \\ +\infty & \text{sinon}, \end{cases}$$

où Φ est une bonne fonction de taux.

La fonction de taux peut être rendue explicite dans certains cas dépendants du support de la “distribution limite” des angles des entrées, ν_1 et ν_2, pour des mesures symétriques. Des encadrements sont aussi disponibles en toute généralité (voir [29, Théorème 1.2] pour plus de détails). Un des traits surprenants de ce PGD est que la fonction de taux met un coût fini seulement sur des mesures qui sont des convolées libres avec la loi du semi-cercle. Biane [23] a montré que la convolution libre avec la loi du semi-cercle est extrêmement régularisante, et donne des mesures absolument continues avec une densité analytique partout où elle est strictement positive. Ainsi, les déviations “autorisées” de la mesure spectrale empirique ne peuvent se faire qu’autour de mesures très régulières. Cette situation contraste avec la fonction de taux des ensembles Gaussiens classiques, pour lesquels, comme nous l’avons fait remarquer, le domaine de la fonction de taux contient des mesures supportées sur n’importe quel compact de dimension de Hausdorff strictement positive. Cette différence traduit des mécanismes de déviations de nature très différentes. On note qu’un analogue de ce théorème de grandes déviations pour la mesure spectrale empirique des matrices de Wishart sans queues Gaussiennes été obtenus par Groux [55], avec une fonction de taux complètement explicite.

Bordenave et Caputo ont montré que les déviations sont dues aux entrées de X qui sont d’ordre $\sqrt{n}$, et que celles-ci forment un réseau sparse qui asymptotiquement devient indépendant du reste de la matrice. Nous allons finir cette section en montrant comment on peut obtenir la borne inférieure de déviations en utilisant seulement une petite proportion des entrées. Soit $\mu = \mu_{sc} \boxplus \nu$ une mesure cible, où ν une mesure symétrique. Soit A une matrice Hermitienne ayant un nombre d’entrées non-nulles négligeable par rapport à n^2 telle que $\mu_A \simeq \nu$. Notons S les indices correspondant aux entrées non-nulles de A. On note aussi $X^{(S)}$ la matrice avec comme entrée (i,j), $X_{i,j}\mathbb{1}_{(i,j)\in S}$, et de façon similaire pour $X^{(S^c)}$. Alors, en utilisant le fait que la convolution libre est uniformément continue [21, Proposition

4.13], et l'inégalité d'Hoeffman-Wielandt (1.2), on peut écrire,

$$\begin{aligned}\mathbb{P}\big(\mu_{X/\sqrt{n}} \simeq \mu_{sc} \boxplus \nu\big) &\gtrsim \mathbb{P}\big(\mu_{X^{(S^c)}/\sqrt{n}+A} \simeq \mu_{sc} \boxplus \mu_A,\ X^{(S)}/\sqrt{n} \simeq A\big)\\ &= \mathbb{P}\big(\mu_{X^{(S^c)}/\sqrt{n}+A} \simeq \mu_{sc} \boxplus \mu_A\big)\mathbb{P}\big(X^{(S)}/\sqrt{n} \simeq A\big).\end{aligned}$$

Par la remarque 1.8.3, on obtient,

$$\mathbb{P}\big(X^{(S)}/\sqrt{n} \simeq A\big) \gtrsim e^{-n^{1+\frac{\alpha}{2}} \frac{1}{n} \sum_{i\leq j} a_{i,j} |A_{i,j}|^{\alpha}}, \tag{1.17}$$

où $a_{i,j} = b$ si $i = j$ et a si $i \neq j$. D'autre part, l'inégalité d'Hoeffman-Wielandt (1.2) implique,

$$\mathcal{W}_2(\mu_{X^{(S^c)}/\sqrt{n}+A}, \mu_{X/\sqrt{n}+A}) \underset{n\to+\infty}{\longrightarrow} 0,$$

en probabilité car $|S| = o(n^2)$. Comme par ailleurs, $X/\sqrt{n}$ et A sont asymptotiquement libres par [3, Théorème 5.4.5], on obtient donc que $\mathbb{P}\big(\mu_{X^{(S^c)}+A/\sqrt{n}} \simeq \mu_{sc} \boxplus \mu_A\big)$ tend vers 1 quand $n \to +\infty$. En choisissant A alternativement diagonale ou diagonale par bloc de taille 2×2, on obtient une borne inférieure de la forme,

$$\mathbb{P}\big(\mu_{X/\sqrt{n}} \simeq \mu_{sc} \boxplus \nu\big) \gtrsim e^{-n^{1+\frac{\alpha}{2}} (b\wedge \frac{a}{2})\nu(|x|^{\alpha})}.$$

1.9 Présentation des résultats

Cette thèse porte sur des problèmes de grandes déviations de spectre de matrices de Wigner et de β-ensembles. Les résultats obtenus reposent principalement sur des phénomènes de queues lourdes, qui permettent, dans les situations que l'on présentera dans la suite, d'obtenir des PGD malgré le fait que la loi jointe du spectre ne sera pas accessible.

1.9.1 Inégalités de concentration

Dans le second chapitre, on s'attachera à obtenir des inégalités de concentration qui reflètent le comportement de grandes déviations de diverses fonctionnelles spectrales, comme la mesure spectrale empirique, la plus grande valeur propre et les traces de polynômes d'une famille de matrices de Wigner. Nous allons dériver ces inégalités de déviations d'une propriété de concentration vérifiée par la loi de la matrice indexée par un paramètre $\alpha \in (0, 2]$ qui rend compte de l'intégrabilité (exponentielle) des entrées. Plus précisément, pour $\alpha \in [1, 2]$, on dira qu'une matrice aléatoire Hermitienne X satisfait la *propriété de concentration* $\mathcal{C}_\alpha$ si pour tout sous-ensemble Borélien $A \subset \mathcal{H}_n^{(\beta)}$, et tout $r > 0$,

$$\mathbb{P}(X \notin A + \sqrt{r}B_{\ell^2} + r^{1/\alpha} B_{\ell^\alpha}) \leq \frac{e^{-Lr}}{\mathbb{P}(X \in A)},$$

pour une certaine constante $L > 0$, où pour tout $p > 0$, on note,

$$B_{\ell^p} = \{H \in \mathcal{H}_n^{(\beta)} : ||H||_{\ell^p} \leq 1\},$$

avec

$$\forall H \in \mathcal{H}_n^{(\beta)},\ ||H||_{\ell^p} = \Big(\sum_{i,j} |H_{i,j}|^p\Big)^{1/p}.$$

Cette définition est motivée par la célèbre inégalité de déviations à deux niveaux de Talagrand [94] pour le produit de lois exponentielles. Le prototype des modèles matriciels satisfaisant $\mathcal{C}_\alpha$ seront pour nous les matrices de Wigner dont les entrées ont une densité proportionnelle à $e^{-c|x|^\alpha}$, pour une certaine constante $c > 0$. On détaillera de nombreux exemples de modèles matriciels vérifiant cette propriété de concentration. Grâce au théorème de Lidskii [22, Corollary III 4.2], on a pour $\alpha \in [1, 2]$, et $A, B \in \mathcal{H}_n^{(\beta)}$,

$$\mathcal{W}_\alpha(\mu_A, \mu_B) \leq \frac{1}{n^{1/\alpha}}||A - B||_{\ell^\alpha}.$$

Cette inégalité nous permet en particulier, de donner une inégalité de concentration pour la mesure spectrale empirique par la Proposition 2.3.1 qui rend compte de la vitesse de grandes déviations des matrices de Wigner sans queues sous-Gaussiennes du Théorème 1.8.4. Nous donnerons aussi une inégalité de déviation pour la plus grande valeur propre par la Proposition 2.5.1 ainsi que pour les traces de polynômes de matrices Hermitiennes qui satisfont $\mathcal{C}_\alpha$ dans la Proposition 2.6.1.

Dans le cas où $\alpha \in (0, 1)$, la propriété de concentration $\mathcal{C}_\alpha$ n'est en réalité plus pertinente. On montre que la mesure de probabilité ν_α^n, où

$$\nu_\alpha = Y_\alpha^{-1} e^{-|x|^\alpha} dx,$$

satisfait, grâce à un argument de transport, l'inégalité de déviation suivante.

1.9.1 Proposition. *Soit $n \in \mathbb{N}$. Il existe une constante $c > 0$ dépendant de α, telle que pour tout $r > 0$, A Borélien de $\mathbb{R}^n$, et $C > 0$ tels que $\nu_\alpha^n(A) > 1/C$,*

$$\nu_\alpha^n\Big(x \notin A + C(\log n)^{\frac{1}{\alpha}-1}\big(\sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) + r^{\frac{1}{\alpha}}B_{\ell^\alpha}\Big) \leq \frac{e^{-cr}}{\nu_\alpha^n(A) - 1/C}.$$

On définit alors une propriété de concentration $\mathcal{C}_\alpha$ pour $\alpha \in (0, 1)$ de même forme que celle vérifiée par ν_α^n. On propose des inégalités de déviations pour la mesure spectrale empirique et la plus grande valeur propre de matrices aléatoires vérifiant cette propriété de concentration, qui reflètent leurs comportement de grandes déviations dans les Propositions 2.3.1 et 2.5.1.

En ce qui concerne la mesure spectrale empirique, obtenir une telle inégalité de concentration nécessite de calculer la constante de Lipschitz de l'application $H \in \mathcal{H}_n^{(\beta)} \mapsto \mu_H$, par rapport à la métrique $|| \ ||_{\ell^\alpha}^\alpha$, au moins pour un certain choix de distance sur $\mathcal{P}(\mathbb{R})$. A cette fin, on définit la distance suivante sur l'ensemble des mesures de probabilité sur $\mathbb{R}$ ayant un moment d'ordre α fini, noté $\mathcal{P}_\alpha(\mathbb{R})$,

$$d_\alpha(\mu, \nu) = \sup_{t \in \mathbb{R}} \Big| \int (t - x)_+^\alpha d\mu(x) - \int (t - x)_+^\alpha d\nu(x) \Big|.$$

En prenant formellement $\alpha = 0$ dans la définition ci-dessus, on retrouve la distance de Kolmogorov-Smirnov. Par une intégration par parties on peut identifier quelle classe de fonctions la distance d_α contrôle, de la même façon que pour la distance de Kolmogorov-Smirnov. En conséquence d'une inégalité due à Rotfel'd [22, Theorem IV.2.14] et Thompson [96], on obtient pour $\alpha \in (0, 1)$, $A, B \in \mathcal{H}_n^{(\beta)}$ et tout $t \in \mathbb{R}$,

$$\Big|\sum_{i=1}^n (t - \lambda_i(A))_+^\alpha - \sum_{i=1}^n (t - \lambda_i(B))_+^\alpha\Big| \leq \sum_{i=1}^n |\lambda_i(A - B)|^\alpha,$$

ce qui permet de calculer la constante de Lipschitz de la mesure spectrale empirique par rapport à $|| \; ||_{\ell^\alpha}^\alpha$ et la distance d_α et d'obtenir une inégalité de déviation pour la mesure spectrale empirique de matrices satisfaisant $\mathcal{C}_\alpha$.

1.9.2 Valeur propre extrême des matrices de Wigner sans queues sous-Gaussiennes

Le troisième chapitre se situe dans la lignée du résultat de Bordenave et Caputo sur les grandes déviations de la mesure spectrale empirique des matrices de Wigner sans queues sous-Gaussiennes. On s'intéressera au problème des grandes déviations de la plus grande valeur propre de ces modèles. Le résultat obtenu est le suivant,

1.9.2 Théorème. *Soit X une matrice de Wigner sans queues sous-Gaussiennes. On suppose de plus que $\Re X_{1,2}$ et $\Im X_{1,2}$ sont indépendantes. La suite $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ suit un PGD de vitesse $n^{\alpha/2}$, et bonne fonction de taux définie pour tout $x \in \mathbb{R}$, par*

$$J_\alpha(x) = \begin{cases} c g_{\mu_{sc}}(x)^{-\alpha} & si\ x > 2, \\ 0 & si\ x = 2, \\ +\infty & si\ x < 2, \end{cases}$$

où c est une constante dépendant de α, a et b, et où $g_{\mu_{sc}}$ désigne la transformée de Stieltjes de la loi du semi-cercle.

La constante c qui apparaît dans ce résultat est solution d'un problème d'optimisation que l'on peut résoudre dans certains cas, en particulier lorsque les entrées sont réelles. La stratégie de la preuve est dans le même esprit que celle développée dans [29]. On utilise des arguments de troncation pour identifier quelles entrées participent aux déviations de la plus grande valeur propre. Puis on se ramène à étudier les déviations de la plus grande valeur propre d'un modèle déformé $H + C$ où H est une matrice dont le spectre est à l'échelle exponentielle essentiellement inclus dans $[-2, 2]$, et C est une matrice ayant seulement une petite proportion d'entrée non-nulles qui sont d'ordre 1. En particulier, cette approche suppose de contrôler la stabilité de l'équation (1.7) vérifiée par la plus grande valeur propre de modèles déformés. On verra que l'on peut en quelque sorte interpréter les déviations de la plus grande valeur propre par des perturbations de rang fini.

1.9.3 Grandes déviations des traces de matrices aléatoires

Dans le chapitre 4, on s'intéresse à un autre exemple de phénomène de queues lourdes que sont les grandes déviations des traces de puissances de matrices aléatoires. On considère trois cas : le cas des β-ensembles associés à un potentiel convexe à croissance polynomial, le cas des matrices de Wigner sous-Gaussiennes, et celui des matrices de Wigner sans queues sous-Gaussiennes. Pour chacun de ces modèles on obtient des PGD pour les moments de la mesure (spectrale) empirique.

Dans le cas des β-ensembles, on a le résultat suivant.

1.9.3 Théorème. *Soient $\alpha \geq 2$ et $\beta > 0$. On suppose que,*

$$\forall x \in \mathbb{R}, \ V(x) = b|x|^\alpha + w(x), \tag{1.18}$$

où w est une fonction convexe telle que $w(x) = o_{\pm\infty}(|x|^\alpha)$. Soit $p \in \mathbb{N}$, $p > \alpha$. Pour tout $\lambda_1, ..., \lambda_n \in \mathbb{R}^n$, on note $m_{p,n}$,

$$m_{p,n} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^p.$$

Sous $\mathbb{P}^n_{V,\beta}$, la suite $(m_{p,n})_{n\geq 1}$ satisfait un PGD à vitesse $n^{1+\frac{\alpha}{p}}$ et de bonne fonction de taux J_p, où $\mathbb{P}^n_{V,\beta}$ est défini par (1.11). Si p est pair,

$$J_p(x) = \begin{cases} b\left(x - \langle \sigma_\beta^V, x^p\rangle\right)^{\frac{\alpha}{p}} & \text{si } x \geq \langle \sigma_\beta^V, x^p\rangle, \\ +\infty & \text{sinon,} \end{cases}$$

où $\langle \sigma_\beta^V, x^p\rangle$ désigne le $p^{\text{ème}}$ moment de la mesure d'équilibre de $\mathbb{P}^N_{V,\beta}$, et si p est impair, J_p est définie par,

$$\forall x \in \mathbb{R},\ J_p(x) = b\left|x - \langle \sigma_\beta^V, x^p\rangle\right|^{\frac{\alpha}{p}}.$$

La stratégie que l'on adoptera pour prouver ce résultat, consiste à montrer d'une part, que les déviations du $p^{\text{ème}}$ moment de la mesure empirique sont dues au $\log n$ particules les plus grandes en valeur absolue. Nous montrerons que la partie du moment $m_{p,n}$ correspondant au $n - \log n$ particules restantes est exponentiellement équivalente à son espérance. Ceci permet de se ramener à comprendre les déviations de moments tronqués où on considère seulement $\log n$ particules, pour lesquelles on pourra montrer que interaction logarithmique est négligeable. La condition $p > \alpha$ nous assure que $m_{p,n}$ n'admet pas de moments exponentiels et donc qu'un phénomène de queues lourdes gouverne les grandes déviations.

En ce qui concerne les matrices de Wigner à entrées Gaussiennes, nous avons le PGD suivant.

1.9.4 Théorème. *Soit $p \in \mathbb{N}$, $p \geq 3$. Soit X une matrice de Wigner centrée à entrées Gaussiennes telle que $\mathbb{E}|X_{1,2}|^2 = 1$. On note τ_n l'état $\frac{1}{n}\mathrm{tr}$ sur $\mathcal{H}_n^{(\beta)}$. La suite $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$, suit un PGD à vitesse $n^{1+\frac{2}{p}}$, et de bonne fonction de taux K_p. Si p est pair, K_p est définie par,*

$$\forall x \in \mathbb{R},\ K_p(x) = \begin{cases} c\left(x - C_{p/2}\right)^{\frac{2}{p}} & \text{si } x \geq C_{p/2}, \\ +\infty & \text{sinon,} \end{cases}$$

où $C_{p/2}$ désigne le $\left(\frac{p}{2}\right)^{\text{ème}}$ nombre de Catalan, et si p est impair,

$$\forall x \in \mathbb{R},\ K_p(x) = c|x|^{\frac{2}{p}}.$$

où c est une constante dépendant de $\mathbb{E}X_{1,1}^2$ et de la structure de covariance de $(\Re X_{1,2}, \Im X_{1,2})$.

Le cas des matrices de Wigner à entrées Gaussiennes a été l'occasion de revisiter la preuve des grandes déviations des chaos de Wiener [69, Section 4] due à Borell et Ledoux. En effet, ce problème peut se reformuler en un problème de grandes déviations de chaos Gaussiens définis sur un espace de dimension croissante. Bien qu'on ne puisse pas déduire des grandes déviations des chaos de Wiener un tel PGD, nous montrerons que le même schéma de preuve peut être mis en œuvre et qu'en un certain sens, les déviations sont dues à des translations. La constante c qui apparaît dans le théorème ci-dessus, peut se calculer explicitement dans le cas où les entrées hors-diagonales sont réelles ou bien si leur covariance est $\frac{1}{2}I_2$.

Enfin, on s'intéressera de nouveau au modèle des matrices de Wigner sans queues sous-Gaussiennes, pour lequel on montrera le résultat de grandes déviations suivant pour les traces de puissances de ces matrices.

1.9.5 Théorème. *Let $p \in \mathbb{N}$, $p \geq 3$. Soit X une matrice de Wigner sans queues sous-Gaussiennes. La suite $(\tau_n(X/\sqrt{n})^p)_{n\geq 1}$ satisfait un PGD à vitesse $n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$ et de bonne fonction de taux J_p. Si p est pair, J_p est donnée par,*

$$\forall x \in \mathbb{R},\ K_{p,\alpha}(x) = \begin{cases} c_p \left(x - C_{p/2}\right)^{\alpha/p} & si\ x \geq C_{p/2}, \\ +\infty & sinon, \end{cases}$$

et si p est impair,

$$\forall x \in \mathbb{R},\ K_{p,\alpha}(x) = c_p|x|^{\alpha/p}.$$

où c_p est une constante dépendant de p, α, a et b.

De plus, si $\alpha \in (0,1]$ et p est pair, alors $c_p = \min\left(b, 2^{-\alpha/p}a\right)$.

On remarque que la fonction de taux correspond essentiellement au coût de déviation d'une entrée $X_{i,j}$ de l'ordre $n^{\frac{1}{2}+\frac{1}{p}}$. En effet, une telle déviation crée une déviation du rayon spectral de l'ordre de $n^{1/p}$, et donc une déviation de la trace normalisée de $(X/\sqrt{n})^p$ de l'ordre de 1. De la même façon que pour les résultats de PGD de la mesure spectrale empirique et de la plus grande valeur propre, les déviations sont expliqués par des grandes entrées de la matrice qui vont ici créer de grandes déviations du bord du spectre.

La similarité des fonctions de taux de ces trois PGD vient du fait que les déviations dans ces trois cas sont gouvernés par le même phénomène de queues lourdes. Dans les trois cas, on observe qu'il n'y a pas de comportement collectif des valeurs propres faisant dévier la trace, les déviations étant seulement dues à des grandes valeurs du spectre.

1.9.4 Une approche isopérimétrique aux grandes déviations

Dans le cinquième chapitre, nous allons revenir sur la preuve de Borell [32], [33] et Ledoux [69] des grandes déviations des chaos de Wiener, dans la continuité du problème des grandes déviations des traces de puissance de matrices de Wigner Gaussiennes. Nous allons montrer que l'approche développée par Borell et Ledoux pour les chaos de Wiener peut servir à expliquer dans une certaine généralité les grandes déviations de modèles où un phénomène de queues lourdes apparaît. Le phénomène marquant qui ressort de la preuve des grandes déviations des chaos de Wiener est que les déviations sont dues à des translations par des éléments de l'espace de Cameron-Martin. Ces mécanismes de déviations sont assez similaires à ceux des déviations du spectre des matrices de Wigner sans queues sous-Gaussiennes, où les déviations sont expliquées par des perturbations additives par des matrices sparses.

Nous proposons un énoncé de grandes déviations général pour une suite de fonctionnelles $f_n : \mathbb{R}^n \to \mathcal{X}$, où $\mathcal{X}$ est un espace métrique, pour lesquelles les grandes déviations, sous la mesure produit ν_α^n, où $\nu_\alpha = Y_\alpha^{-1} e^{-|x|^\alpha} dx$, sont créées par des translations. Comme application de ce résultat nous retrouverons les PGD connus pour le spectre des matrices de Wigner sans queues sous-Gaussiennes. Pour des matrices de Wigner ayant des entrées de partie réelle et imaginaire indépendantes de densité proportionnelle à $e^{-c|x|^\alpha}$ pour $c > 0$ et $\alpha \in (0, 2)$, les PGD pour la mesure spectrale empirique, la plus grande valeur propre et les moments de la mesure spectrale empirique se déduiront d'une manière unifiée de ce résultat plus général de grandes déviations.

1.9.6 Théorème. *Soit $(\mathcal{X}, d)$ un espace métrique. Soient $\alpha \in (0, 2]$ et $N \subset \mathbb{N}$ une partie infinie. Soit X_n une variable aléatoire distribuée suivant la loi ν_α^n. Soit $f_n, F_n : \mathbb{R}^n \to \mathcal{X}$ une suite d'applications mesurables. Soit $(v(n))_{n\in N}$ une suite qui tend vers $+\infty$. On définit pour $\delta > 0$ et $n \in N$, la fonction,*

$$\forall x \in \mathcal{X},\ I_{n,\delta}(x) = \inf\{||h||_{\ell^\alpha}^\alpha : d(F_n(h), x) < \delta, h \in \mathbb{R}^n\}.$$

On pose,

$$\forall x \in \mathcal{X},\ I_\alpha(x) = \sup_{\delta>0} \inf_{n\in N} I_{n,\delta}(x). \tag{1.19}$$

On suppose:
(i). (Équivalent déterministe uniforme). Pour tout $r > 0$,

$$\sup_{h_n \in r B_{\ell^\alpha}} d\big(f_n(X_n + v(n)^{1/\alpha} h_n), F_n(h_n)\big) \underset{\substack{n\to+\infty \\ n\in N}}{\longrightarrow} 0,$$

en probabilité.
(ii). (Contrôle de la constante de Lipschitz). Si $\alpha < 2$, alors pour tout $\delta > 0$ et $r > 0$, il existe une suite $t_\delta(n)$ telle que,

$$\mathbb{E} \sup_{||h||_2 \le t_\delta(n)} \mathcal{L}_n(h) \le \delta,$$

avec

$$\mathcal{L}_n(h) = \sup_{X_n + r v(n)^{1/\alpha} B_{\ell^\alpha}} d\big(f_n(x+h), f_n(x)\big),$$

satisfaisant

$$(\log n)^{\alpha/2} = o(\log \frac{t_\delta(n)^2}{v(n)})\ \textit{si}\ \alpha \neq 1,\ \ \textit{ou}\ v(n) = o(t_\delta(n)^2)\ \textit{si}\ \alpha = 1.$$

(iii). (Compacité). Pour tout $r > 0$, $\cup_{n\in N} F_n(r B_{\ell^\alpha})$ est relativement compact.
(iv). (Borne supérieure = borne inférieure). Pour tout $x \in \mathcal{X}$,

$$I_\alpha(x) = \sup_{\delta>0} \limsup_{\substack{n\to+\infty \\ n\in N}} I_{n,\delta}(x).$$

Alors $(f_n(X_n))_{n\in N}$ satisfait un PGD à vitesse $v(n)$ et de bonne fonction de taux I_α.

Notons $\mathcal{S}_\alpha$ l'ensemble des matrices de Wigner ayant pour densité $Z_{W_\alpha}^{-1}e^{-W_\alpha}$ par rapport à la mesure de Lebesgue sur $\mathcal{H}_n^{(\beta)}$, notée $\ell_n^{(\beta)}$, où W_α est de la forme,

$$\forall H \in \mathcal{H}_n^{(\beta)},\ W_\alpha(H) = b\sum_{i=1}^n |H_{i,i}|^\alpha + \sum_{i<j}\big(a_1|\Re X_{i,j}|^\alpha + a_2|\Im X_{i,j}|^\alpha\big). \tag{1.20}$$

En conséquence du résultat de grandes déviations du Théorème 1.9.6, on donnera une extension des résultats du chapitre 4, on montrant le PGD suivant pour les traces de polynômes d'une famille de matrices de Wigner indépendantes dans la classe $\mathcal{S}_\alpha$ pour $\alpha \in (0,2]$.

1.9.7 Théorème. *Soient $\alpha \in (0,2]$ et $p \in \mathbb{N}$, $p > \alpha$. Soit $\boldsymbol{X} = (X_1, ..., X_p)$ est une famille de matrices de Wigner indépendantes dans la classe $\mathcal{S}_\alpha$, telles que pour tout $M \in \{X_1, ..., X_p\}$, $\mathbb{E}|M_{1,2}|^2 = 1$. On suppose que X_i a pour loi $Z_{W_\alpha}^{-1}e^{-W_{\alpha,i}}d\ell_n^{(\beta)}$, où $W_{\alpha,i}$ est de la forme (1.20). Soit $P \in \mathbb{C}\langle \boldsymbol{X}\rangle$ un polynôme non-commutatif de degré total d. On note τ_n l'état $\frac{1}{n}\mathrm{tr}$ sur $\mathcal{H}_n^{(\beta)}$. La suite*

$$\tau_n[P(\boldsymbol{X}/\sqrt{n})]$$

satisfait un PGD à vitesse $n^{\alpha\left(\frac{1}{2}+\frac{1}{d}\right)}$ et de bonne fonction de taux K_α, définie pour tout $x \in \mathbb{R}$ par,

$$K_\alpha(x) = \begin{cases} c_1\big(x - \tau[P(\boldsymbol{s})]\big)^{\frac{\alpha}{d}} & \text{si } x > \tau[P(\boldsymbol{s})], \\ 0 & \text{si } x = \tau[P(\boldsymbol{s})], \\ c_{-1}\big|x - \tau[P(\boldsymbol{s})]\big|^{\frac{\alpha}{d}} & \text{si } x < \tau[P(\boldsymbol{s})], \end{cases}$$

où pour $\sigma \in \{-1, 1\}$,

$$c_\sigma = \inf\big\{W_\alpha(\boldsymbol{H}) : \boldsymbol{H} \in \cup_{n\in\mathbb{N}}(\mathcal{H}_n^{(\beta)})^p,\ \sigma = \mathrm{tr}P_d(\boldsymbol{H})\big\} \in [0, +\infty],$$

où $W_\alpha(\boldsymbol{H}) = \sum_{i=1}^p W_{\alpha,i}(H_i)$, P_d est la partie homogène de degré d de P, et où $\boldsymbol{s} = (s_1, ..., s_p)$ est une famille libre de p variables semi-circulaires sur un espace de probabilité non-commutatif $(\mathcal{A}, \tau)$. De plus, si d est impair $c_1 = c_{-1}$.

Enfin, nous présentons une application du Théorème 1.9.6 au temps de dernier passage dans le cas où les poids sont distribués suivant la mesure de probabilité $\mu_\alpha = Z_\alpha^{-1}e^{-x^\alpha}\mathbb{1}_{x\geq 0}dx$, où $\alpha \in (0,1)$. Soit $d \geq 2$. On définit le temps de dernier passage $T(X)$, par

$$T(X) = \sup_\pi \sum_{v\in\pi} X_v,$$

où le supremum porte sur les chemins dirigés dans $\mathbb{Z}_+^d$, de $(0, ..., 0)$ à $(n, ..., n)$, où chaque pas consiste à augmenter une coordonnée de 1. On sait grâce à un résultat de Martin [75], que si les poids X_v sont i.i.d de fonction de répartition commune F satisfaisant,

$$\int_0^{+\infty}(1 - F(t))^{1/d}dt < +\infty,$$

alors,

$$\frac{1}{n}\mathbb{E}T(X) \underset{n\to+\infty}{\longrightarrow} g(1, ..., 1),$$

où $g(1, ..., 1)$ est déterministe. Avec ces notations, on obtient comme application du Théorème 1.9.6 le résultat suivant.

1.9.8 Théorème. *Soit $\alpha \in (0,1)$. Soit $(X_v)_{v\in\mathbb{Z}_+^d}$ une famille de variables i.i.d distribuées selon μ_α. La suite $T(X)/n$ satisfait un PGD à vitesse n^α et de bonne fonction de taux L_α, définie par,*

$$L_\alpha(x) = \begin{cases} (x - g(1, ..., 1))^\alpha & \text{if } x \geq g(1, ..., 1), \\ +\infty & \text{otherwise.} \end{cases}$$

1.9.5 Grandes déviations de la mesure empirique des β-ensembles

Enfin, dans un dernier chapitre, nous allons revisiter le PGD de la mesure spectrale empirique des β-ensembles du Théorème 1.8.1. Dans le cas où le potentiel est quadratique, nous allons utiliser la représentation matricielle tridiagonale pour donner une nouvelle preuve à ce résultat de grandes déviations. En particulier, dans le cas du GOE ou du GUE, cette approche fournit une preuve qui ne repose pas sur la connaissance de la loi du spectre. La stratégie de preuve repose sur la contraction d'un PGD que l'on obtient pour l'opérateur de Schrödinger aléatoire donné par tridiagonalisation pour la topologie que l'on a évoqué au paragraphe 1.7.3.

Comme produit dérivé de cette preuve, on obtient une formule variationnelle pour l'entropie non-commutative d'une mesure, dans le même esprit que les travaux de Gamboa, Nagel et Rouault sur les "sum rules" [52].

1.9.9 Proposition. *On note $\mathcal{P}_{stat}$ l'ensemble des mesures de probabilité stationnaires sur $\mathbb{R}^\mathbb{Z} \times \mathbb{R}_+^\mathbb{Z}$. Soit $\mu \in \mathcal{P}(\mathbb{R})$ telle que $\mu(x^2) < +\infty$. On a l'égalité suivante dans $\mathbb{R} \cup \{-\infty\}$,*

$$\int \log|x-y| d\mu(x)d\mu(y) = 2\sup_\pi \mathbb{E}_\pi T \log w_0,$$

où le supremum porte sur tous les couplages π entre $\rho \in \mathcal{P}_{stat}$ tel que $\mu_\rho = \mu$, et la loi uniforme sur $[0,1]$, et où sous $\mathbb{P}_\pi$, $((v_k, w_k)_{k\in\mathbb{Z}}, T)$ a pour loi π.

2. Concentration inequalities

2.1 Introduction

We will work in this chapter at deriving concentration inequalities for some spectral functionals of Wigner matrices. The stake for us behind these inequalities is to capture the large deviation behavior of those functionals. They will play a major role in our understanding in a first phase of the speed of the deviations, and in a second phase of the mechanism of deviations. As it will happens, some of the concentration inequalities we provide will also enable us to capture the fluctuations of our functionals.

To derive such concentration inequalities for functions of the spectrum of random matrices, we will follow the classical argument which consists in considering our functionals as functions of the entries, and taking advantage of the concentration property of the law of the underlying random matrix. This approach is made possible in the setting where the spectrum is a smooth function of the entries, which will be our case as we will work with Hermitian matrices.

For Hermitian random matrices with bounded entries or with entries satisfying a Log-Sobolev inequality, concentration inequalities for Lipschitz (convex) linear statistics of the eigenvalues or for the largest eigenvalue, have been extensively studied by Guionnet-Zeitouni [57], Guionnet [56, Part II] and Ledoux [70, Chapter 8 §8.5].

We will provide here some concentration inequalities for certain functions of the spectrum under weaker assumptions on the concentration property of the law of the matrix, than normal concentration. More precisely, we will provide concentration inequalities for the spectral measure, the largest eigenvalue, and traces of polynomials of random Hermitian matrices satisfying a certain concentration property which will be indexed by some $\alpha \in (0, 2]$. As will see, this concentration property will entail a grading of speeds of deviations for the spectral functionals we are interested in, as it has been observed in Theorem 1.8.4 for the deviations of the empirical spectral measure of Wigner matrices without Gaussian tails.

2.2 Concentration property $\mathcal{C}_\alpha$

We now present the concentration property with which we will be working. Let $\alpha \in [1,2]$. We will say in the following that a random Hermitian matrix X satisfies the *concentration property* $\mathcal{C}_\alpha$, if there is a constant $\kappa > 0$, such that for any Borel subset A of $\mathcal{H}_n^{(\beta)}$, such that $\mathbb{P}(X \in A) \geq 1/2$, and any $t > 0$,

$$\mathbb{P}(X \notin A + \kappa\sqrt{t}B_{\ell^2} + \kappa t^{1/\alpha}B_{\ell^\alpha}) \leq 2e^{-t}, \tag{2.1}$$

where for any $p > 0$,

$$B_{\ell^p} = \{Y \in \mathcal{H}_n^{(\beta)} : ||Y||_{\ell^p} \leq 1\},$$

with

$$\forall Y \in \mathcal{H}_n^{\beta},\ ||Y||_{\ell^p}^p = \sum_{i,j} |Y_{i,j}|^p.$$

The concentration property $\mathcal{C}_\alpha$ is equivalent (see [70, Proposition 1.3]) to the following deviation inequality of Lipschitz functions around their medians, which will be useful in the applications.

2.2.1 Lemma. *Let X be a random Hermitian matrix satisfying $\mathcal{C}_\alpha$ for some $\kappa > 0$. Let $f : \mathcal{H}_n^{(\beta)} \to \mathbb{R}$ be a function respectively L_2-Lipschitz and L_α-Lipschitz with respect to $||\ ||_{\ell^2}$, and $||\ ||_{\ell^\alpha}$. Then, for any $t > 0$,*

$$\mathbb{P}(f(X) > m_f + t) \leq 2\exp\Big(-\min\Big(\frac{t^2}{4\kappa L_2^2}, \frac{t^\alpha}{2^\alpha \kappa L_\alpha^\alpha}\Big)\Big),$$

where m_f denotes the median of $f(X)$.

In the following we set ν_α, respectively μ_α to be the probability measure on $\mathbb{R}$, respectively $\mathbb{R}_+$, with density $Y_\alpha^{-1}e^{-|x|^\alpha}$, respectively $Z_\alpha^{-1}e^{-x^\alpha}$.

The reason for defining this concentration property comes from Talagrand's famous two-levels deviation inequality for the exponential law [94], which says that for any Borel subset A of $\mathbb{R}^n$ with $\nu_1^n(A) > 0$, and $r > 0$,

$$\nu_1^n(x \notin A + \sqrt{r}B_{\ell^2} + rB_{\ell^1}) \leq \frac{e^{-Lr}}{\nu_1^n(A)}, \tag{2.2}$$

for some $L > 0$, and a similar inequality for μ_1^n. Using a transport argument, one can translate (2.2) for the probability measures ν_α^n, and μ_α^n for $\alpha \geq 1$, and deduce the following deviation inequality,

$$\nu_\alpha^n(x \notin A + \sqrt{r}B_{\ell^2} + r^{\frac{1}{\alpha}}B_{\ell^\alpha}) \leq \frac{e^{-L'r}}{\nu_\alpha^n(A)}.$$

We will briefly review some examples of random Hermitian matrices having concentration $\mathcal{C}_\alpha$. We will associate to the three regimes $\alpha = 2$, $\alpha = 1$ and $\alpha \in (1,2)$ a functional inequality which will provide a necessary condition to the concentration property $\mathcal{C}_\alpha$. We start with the better known case $\alpha = 2$ of measures with *normal concentration* .

2.2.1 Log-Sobolev inequalities

Let μ be a probability measure on $\mathbb{R}^n$. We say that μ satisfies a *Log-Sobolev inequality* with constant $c > 0$ if for any smooth function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Ent}_\mu(f^2) \leq c \int |\nabla f|^2 d\mu,$$

where $\mathrm{Ent}_\mu(f^2)$ is the entropy of f^2 which is defined by

$$\mathrm{Ent}_\mu(f^2) = \int f^2 \log f^2 d\mu - \Big(\int f^2 d\mu \Big) \Big(\int \log f^2 d\mu \Big),$$

if $\int f \log(1+f^2) d\mu < +\infty$, and by $+\infty$ otherwise.

The first example of probability measure which satisfies a Log-Sobolev inequality is the Gaussian measure by [70, Theorem 5.1]. The probability measures ν_α with density $Y_\alpha^{-1} e^{-|x|^\alpha}$ with respect to Lebesgue measure, also satisfy a Log-Sobolev inequality when $\alpha \geq 2$ [13, Chapter 7 §7.7]. Log-Sobolev inequalities enjoy many agreeable properties as a *tenzorisation property* and a *stability property* under perturbation by a bounded potential.

Stability. If μ satisfies a Log-Sobolev inequality with constant $c > 0$, and V is a bounded potential, then $Z^{-1} e^{V} \mu$, where Z is the normalization factor, satisfies a Log-Sobolev inequality with constant $ce^{4||V||_\infty}$ (see [70, Proposition 5.5]).

Tensorization. If μ_i, satisfies a Log-Sobolev inequality with constant c_i for $i = 1, 2$, then $\mu_1 \otimes \mu_2$ satisfies a Log-Sobolev inequality with constant $\max(c_1, c_2)$ (see [70, Corollary 5.7]).

Log-Sobolev inequalities entails *normal concentration* by the Herbst argument (see [70, Theorem 5.3]). More precisely, if μ satisfies a Log-Sobolev with constant $c > 0$, then for any 1-Lispchitz and integrable function f,

$$\mu\Big(f - \int f d\mu > t\Big) \leq e^{-t^2/4c},$$

which implies by [70, Proposition 1.7] that for any Borel subsets A of $\mathbb{R}^n$ with $\mu(A) \geq 1/2$, and any $r > 0$,

$$\mu\big(x \notin A + \sqrt{r} B_{\ell^2}\big) \leq e^{-r^2/8c}.$$

We readily deduce from the tensorization property of the Log-Sobolev inequality, that the law of any matrix in the class $\mathcal{S}_\alpha$, that is, with entries whose density is proportional to $e^{-c|x|^\alpha}$ for some constant $c > 0$, satisfies a Log-Sobolev inequality with a constant independent of the dimension, when $\alpha \geq 2$.

It is known by that any probability measure $\mu = Z^{-1} e^{-V} dx$ on $\mathbb{R}^n$ where V is a potential uniformly strictly convex, in the sense that $\mathrm{Hess}(V)(x) \geq c > 0$, satisfies by [70, Theorem 5.2], a Log-Sobolev inequality with constant $2/c$.

Thus, if X is a Wigner matrix with entries whose law are uniformly strictly log-concave, then the law of X satisfies a Log-Sobolev inequality with a constant independent of the dimension, and thus, the concentration property $\mathcal{C}_2$. Another important example of random Hermitian matrices are unitarily invariant random matrix models which are sampled under the law with density proportional to $e^{-\mathrm{tr}V}$ with respect to the Lebesgue measure on $\mathcal{H}_n^{(\beta)}$. By Klein's lemma (see [3, Lemma 4.4.12]), we deduce that if V is a uniformly strictly convex function, then the probability measure $e^{-\mathrm{tr}V}d\ell_n^{(\beta)}$ satisfies a Log-Sobolev inequality.

2.2.2 Poincaré inequality

Let μ be a probability measure on $\mathbb{R}^m$. We say that μ satisfies a *Poincaré inequality* if there is a $\lambda > 0$ such that for any smooth $f : \mathbb{R}^m \to \mathbb{R}$,

$$\lambda \mathrm{Var}_\mu f \leq \int |\nabla f|^2 d\mu,$$

where $|\ |$ is the Euclidean norm on $\mathbb{R}^m$. The constant λ appearing in the Poincaré inequality is called the *spectral gap*. The first example of probability measure satisfying such an inequality is the exponential measure, or the symmetric exponential measure. As the Log-Sobolev inequality, Poincaré inequality tensorizes and entail a Talagrand-type concentration property.

Tensorization. By the Efron-Stein inequality [76, Theorem 3.1], we see that if μ_i satisfies the Poincaré inequality with spectral gap λ_i, for $i = 1, 2$, then $\mu_1 \otimes \mu_2$ satisfies Poincaré inequality with spectral gap $\min(\lambda_1, \lambda_2)$.

Talagrand-type concentration. By Bobkov-Ledoux [25, Corollary 3.2], we know that if μ satisfies the Poincaré inequality, then one has the following deviation inequality for the product measure μ^n. For any Borel subset of $(\mathbb{R}^m)^n$ such that $\mu^n(A) > 0$, and any $r > 0$,

$$\mu^n\big(x \notin A + \sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) \leq \frac{e^{-Lr}}{\mu^n(A)},$$

for some L depending on the spectral gap.

Bobkov [26] showed that any log-concave law on $\mathbb{R}^n$ satisfies a Poincaré inequality with a certain spectral gap depending on the dimension. Thus, any Wigner matrix with entries whose laws are log-concave will satisfy $\mathcal{C}_1$.

2.2.3 Poincaré-type inequality

To reach the intermediate regime of random matrices having concentration $\mathcal{C}_\alpha$, we introduce the notion of *Poincaré-type inequalities*. Let d_m be some distance on $\mathbb{R}^m$. For a smooth function $f : \mathbb{R}^m \to \mathbb{R}$, we define the length of the gradient of f with respect to the distance d_m by,

$$\forall x \in \mathbb{R}^m,\ |\nabla f(x)| = \limsup_{y \to x} \frac{|f(y) - f(x)|}{d_m(x, y)}.$$

We say that a probability measure μ satisfies a *Poincaré-type inequality* on $(\mathbb{R}^m, d_m)$ if there is some $\lambda > 0$, such that for any smooth $f : \mathbb{R}^m \to \mathbb{R}$,

$$\lambda \mathrm{Var}_\mu f \leq \int |\nabla f|^2 d\mu,$$

where the length of the gradient is taken with respect to d_m.

Let $\alpha \in (1,2)$ and define $\omega(t) = \mathrm{sg}(t)\max(|t|,|t|^\alpha)$ for any $t \in \mathbb{R}$. We denote by d the distance on $\mathbb{R}$ defined by

$$\forall x, y \in \mathbb{R},\ d(x,y) = |\omega(x) - \omega(y)|,$$

and by d_n the distance on $\mathbb{R}^n$, $d_n(x,y) = \big(\sum_{i=1}^n d(x_i,y_i)^2\big)^{1/2}$.

Bobkov-Ledoux [25, Corollary 3.2], and Gozlan [54, Proposition 2.4], proved that the Poincaré-type inequality on $(\mathbb{R}^m, d_m)$ yields a certain concentration property for the product measure of the following form: if μ satisfies a Poincaré-type inequality on $(\mathbb{R}^m, d_m)$, then there is some $L > 0$ such that for any $n \in \mathbb{N}$, any Borel subset A of $(\mathbb{R}^m)^n$ with $\mu^n(A) > 0$, and any $r > 0$,

$$\mu^n(x \notin A + \sqrt{r}B_{\ell^2} + r^{\frac{1}{\alpha}}B_{\ell^\alpha}) \leq \frac{e^{-Lr}}{\mu^n(A)}.$$

Due to Gozlan [54, Proposition 3.3], any probability measure μ on $\mathbb{R}$ with density $Z_V^{-1}e^{-V}$ with a C^1 potential V verifying,

$$\liminf_{x\to\pm\infty} \frac{\mathrm{sg}(x)V'(x)}{x^{\alpha-1}} > 0, \tag{2.3}$$

satisfies a Poincaré-type inequality on $(\mathbb{R}, d)$. We deduce that Wigner matrices with entries whose laws have a density $Z_V^{-1}e^{-V}$, with V verifying the criterion (2.3), have concentration $\mathcal{C}_\alpha$.

We saw that the Wigner matrices in the class $\mathcal{S}_\alpha$ are somehow the prototypes of Wigner matrices having concentration $\mathcal{C}_{\alpha\wedge 2}$ for $\alpha \geq 1$. But what can be said when $\alpha < 1$?

2.2.4 A deviation inequality for ν_α^n, $\alpha \in (0,1)$

We know by Talagrand [93] that as ν_α does not have an exponential tail, ν_α^n cannot satisfy a deviation inequality which does not depend on the dimension. It can be shown that the probability measure ν_α^n satisfies a weak Poincaré inequality (see [13, Chapter 7 §7.5]). The derivation of a deviations inequality from the weak Poincaré inequality has been investigated by Barthe, Cattiaux and Roberto [14], and yields a concentration inequality with respect to Euclidean enlargements. We will follow another path which consists, as for the case $\alpha \geq 1$, in transporting Talagrand's deviation inequality (2.2) for ν_1 to ν_α with $\alpha < 1$. We start with the one-sided probability measure μ_α.

2.2.2 Proposition. *Let $n \in \mathbb{N}$, $n \geq 2$, and $\alpha \in (0,1)$. There is a constant $c > 0$ depending on α, such that for any $r > 0$, A a measurable subset of $\mathbb{R}_+^n$, and $C > 0$ such that $\mu_\alpha^n(A) > 1/C$,*

$$\mu_\alpha^n\Big(x \notin A + C(\log n)^{\frac{1}{\alpha}-1}\big(\sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) + r^{\frac{1}{\alpha}}B_{\ell^\alpha}\Big) \leq \frac{e^{-cr}}{\mu_\alpha^n(A) - 1/C}.$$

2.2.3 Remark. This deviation inequality is not optimal in the sense that it fails to capture the Gaussian fluctuations of empirical means from the central limit theorem. This is due to the $(\log n)^{1/\alpha-1}$ factor in front of the ℓ^2-ball, which comes from the fact that the increasing rearrangement from μ_1 to μ_α is not a Lipschitz function.

But on the other hand, for larger deviations, the $(\log n)^{\frac{1}{\alpha}-1}$ factor seems to be sharp, since it yields a non-trivial deviation inequality for

$$(\log n)^{\frac{1}{\alpha}-1}\big(\max_{1\leq i\leq n} x_i - m\big),$$

where m is the median of the maximum function under μ_α^n. But from the extreme theory (see [67, Theorem 1.6.2, Corollary 1.6.3]),

$$a_n\big(\max_{1\leq i\leq n} x_i - b_n\big),$$

converges in law to the Gumbel distribution G, where

$$a_n \sim c_1(\log n)^{\frac{1}{\alpha}-1}, \text{ and } b_n \sim c_2(\log n)^{\frac{1}{\alpha}},$$

for some constant c_1, c_2. Moreover, as the Gumbel distribution has a right-tail behaving like e^{-t}, we see that the B_{ℓ^1} part in the enlargement of the deviations inequality of Proposition 2.2.2 is justified.

If μ and ν are two probability measures on $\mathbb{R}$, we define the monotone rearrangement T of μ onto ν by,

$$\forall t \in \mathbb{R},\ \mu(-\infty, t] = \nu(-\infty, \varphi(t)].$$

This defines a unique non-decreasing map if the distribution function of ν is invertible which sends μ to ν. To transport (2.2), we will need an estimate on the monotone rearrangement φ of μ_1 onto μ_α, which is given by the following lemma.

2.2.4 Lemma. *Let $\alpha \in (0,1)$. Let φ be the monotone rearrangement of μ_1 onto μ_α. There is a constant $K > 0$ depending on α such that for any $x, y \in [0, +\infty)$,*

$$|\varphi(x) - \varphi(y)| \leq K \max\Big(|x-y|, x^{\frac{1}{\alpha}-1}|x-y|, |x-y|^{\frac{1}{\alpha}}\Big).$$

Proof. This proof is very much in the spirit of [94, Lemma 2.5]. We begin by bounding from above

$$\int_x^{+\infty} e^{-y^\alpha} dy,$$

when $x \geq 1$. The change of variable $u = y^\alpha$ gives,

$$\int_x^{+\infty} e^{-y^\alpha} dy = \frac{1}{\alpha}\int_{x^\alpha}^{+\infty} u^{\frac{1}{\alpha}-1} e^{-u} du.$$

Let $m = \lceil \frac{1}{\alpha} \rceil$. Integrating by parts m times, we get

$$\begin{aligned}\int_{x^\alpha}^{+\infty} u^{\frac{1}{\alpha}-1} e^{-u} du = &\sum_{k=1}^{m-1} \big(\frac{1}{\alpha}-1\big)\ldots\big(\frac{1}{\alpha}-k+1\big) x^{1-k\alpha} e^{-x^\alpha} \\ &+ \big(\frac{1}{\alpha}-1\big)\ldots\big(\frac{1}{\alpha}-m+1\big)\int_{x^\alpha}^{+\infty} u^{\frac{1}{\alpha}-m} e^{-u} du.\end{aligned}$$

As $x \geq 1$, and $\frac{1}{\alpha} - m \leq 0$, we deduce

$$\int_{x^\alpha}^{+\infty} u^{\frac{1}{\alpha}-1} e^{-u} du \leq K x^{1-\alpha} e^{-x^\alpha},$$

where $K > 0$ is some constant depending on α which will vary along the proof. Therefore, for any $x \geq 1$,

$$\int_x^{+\infty} e^{-y^\alpha} dy \leq K x^{1-\alpha} e^{-x^\alpha}. \tag{2.4}$$

By definition φ satisfies for any $x > 0$,

$$e^{-x} = \int_{\varphi(x)}^{+\infty} e^{-y^\alpha} \frac{dy}{Z_\alpha}. \tag{2.5}$$

This implies that φ is an increasing homeomorphism of $\mathbb{R}_+$. For $\varphi(x) \geq 1$, we have

$$e^{-x} \leq K \varphi(x)^{1-\alpha} e^{-\varphi(x)^\alpha}. \tag{2.6}$$

From (2.5), we see that φ is differentiable, and φ' satisfies for any $x \geq 0$,

$$e^{-x} = \frac{1}{Z_\alpha} \varphi'(x) e^{-\varphi(x)^\alpha}.$$

Thus by (2.6), we get for $t \geq \varphi^{-1}(1)$,

$$\varphi'(t) \leq K \varphi(t)^{1-\alpha}. \tag{2.7}$$

Dividing by $\varphi(t)^{1-\alpha}$ and integrating on $[\varphi^{-1}(1), x]$ we get

$$\varphi(x)^\alpha - 1 \leq K(x - \varphi^{-1}(1)),$$

for any $x \geq \varphi^{-1}(1)$. Hence,

$$\varphi(x) \leq K x^{\frac{1}{\alpha}}, \tag{2.8}$$

for $x \geq \varphi^{-1}(1)$. By (2.7) we deduce

$$\varphi'(x) \leq K x^{\frac{1}{\alpha}-1}.$$

Since φ' is continuous, at the price of taking K larger, we have

$$\forall x \geq 0,\ \ \varphi'(x) \leq K \max(1, x^{\frac{1}{\alpha}-1}).$$

Let $x \geq 0$, and $y \in \mathbb{R}$ such that $x + y \geq 0$. If $x, x+y \leq 1$,

$$|\varphi(x+y) - \varphi(x)| \leq K y.$$

Whereas if $x, x+y \geq 1$,

$$|\varphi(x+y) - \varphi(x)| \leq K \int_x^{x+y} t^{\frac{1}{\alpha}-1} dt = \alpha K\big((x+y)^{\frac{1}{\alpha}} - x^{\frac{1}{\alpha}}\big).$$

Now, if $0 \leq x \leq 1 \leq x+y$,

$$\begin{aligned}|\varphi(x+y)-\varphi(x)| &\leq K\int_x^{x+y}(1+t^{\frac{1}{\alpha}-1})dt\\ &\leq K\big(y+\alpha\big((x+y)^{\frac{1}{\alpha}}-x^{\frac{1}{\alpha}}\big).\end{aligned}$$

In conclusion, for any $x \geq 0$, $x+y \geq 0$,

$$|\varphi(x+y)-\varphi(x)| \leq K\max\Big(y, \big((x+y)^{\frac{1}{\alpha}}-x^{\frac{1}{\alpha}}\big)\Big). \tag{2.9}$$

The mean value theorem yields

$$|(x+y)^{\frac{1}{\alpha}}-x^{\frac{1}{\alpha}}| \leq \frac{1}{\alpha}\max\big(x^{\frac{1}{\alpha}-1},(x+y)^{\frac{1}{\alpha}-1}\big)y.$$

Using the convexity of $x \mapsto |x|^{\frac{1}{\alpha}-1}$, if $1/\alpha \geq 1$, or its sub-additivity, when $1/\alpha - 1 \in (0,1)$, we get

$$|(x+y)^{\frac{1}{\alpha}}-x^{\frac{1}{\alpha}}| \leq \frac{a_\alpha}{\alpha}\max\big(x^{\frac{1}{\alpha}-1},x^{\frac{1}{\alpha}-1}+y^{\frac{1}{\alpha}-1}\big)y,$$

with $a_\alpha = \max(1, 2^{\frac{1}{\alpha}-2})$. Together with (2.9), this gives the claim. □

With this estimate on the monotone rearranegement, we are now ready to prove Proposition 2.2.2.

Proof of Proposition 2.2.2. Let $\Phi = \varphi^{\otimes n} : \mathbb{R}^n \to \mathbb{R}^n$, defined by $\Phi(x) = (\varphi(x_i))_{1\leq i\leq n}$, which sends μ_1^n to μ_α^n. Let $r > 0$, and A be a measurable subset of $\mathbb{R}_+^n$ such that $\mu_1^n(A) > 0$. In a first step, we will use Lemma 2.2.4 to see how the map Φ transform the set $A + \sqrt{r}B_{\ell^2} + rB_{\ell^1}$. Actually, to transport the deviation inequality of μ_1^n it is sufficient to understand how Φ deforms $A' + \sqrt{r}B_{\ell^2} + rB_{\ell^1}$ for a well-chosen subset A' of A such that $\mu_1^n(A') > 0$. Define

$$B = \{x \in \mathbb{R}^n : ||x||_\infty \leq C\log n\},\ \ A' = A \cap B,$$

where C is some constant which will be chosen later. Let $x \in A'$, $y \in B_{\ell^2}$, and $z \in B_{\ell^1}$. By Lemma 2.2.4, we have

$$|\Phi(x+\sqrt{r}y)-\Phi(x)| \leq K\big(\sqrt{r}|y| + |x|^{\frac{1}{\alpha}-1}\sqrt{r}|y| + |\sqrt{r}y|^{\frac{1}{\alpha}}\big),$$

where the inequality has to be understood coordinate-wise, and the functions are applied coordinate by coordinate to the vectors in $\mathbb{R}^n$, and where K is a constant depending on α which will vary in the rest of the proof without changing name. Thus,

$$\Phi(x+\sqrt{r}y)-\Phi(x) \in K\Big(\sqrt{r}B_{\ell^2} + (C\log n)^{\frac{1}{\alpha}-1}\sqrt{r}B_{\ell^2} + r^{\frac{1}{2\alpha}}B_{\ell^{2\alpha}}\Big).$$

For $C\log n \geq 1$, we have

$$\Phi(x+\sqrt{r}y)-\Phi(x) \in K\Big((C\log n)^{\frac{1}{\alpha}-1}\sqrt{r}B_{\ell^2} + r^{\frac{1}{2\alpha}}B_{\ell^{2\alpha}}\Big).$$

Once again by Lemma 2.2.4, we get

$$|\Phi(x+\sqrt{r}y+rz)-\Phi(x+\sqrt{r}y)| \le K\big(|rz|+|x+\sqrt{r}y|^{\frac{1}{\alpha}-1}|rz|+|rz|^{\frac{1}{\alpha}}\big),$$

where again this inequality is valid coordinate-wise. Using the convexity of the power function $t \mapsto |t|^{\frac{1}{\alpha}-1}$, or its sub-additivity, we get

$$|\Phi(x+\sqrt{r}y+rz)-\Phi(x+\sqrt{r}y)| \le K\big(|rz|+(|x|^{\frac{1}{\alpha}-1}+|\sqrt{r}y|^{\frac{1}{\alpha}-1})|rz|+|rz|^{\frac{1}{\alpha}}\big).$$

Note that Hölder's inequality implies

$$|y|^{\frac{1}{\alpha}-1}|z| \in B_{\ell^\gamma},$$

with $\frac{1}{\gamma}=\frac{1}{2}(\frac{1}{\alpha}+1)$. Thus,

$$\Phi(x+\sqrt{r}y+rz)-\Phi(x+\sqrt{r}y) \in K\big((C\log n)^{\frac{1}{\alpha}-1}rB_{\ell^1}+r^{\frac{1}{\gamma}}B_{\ell^\gamma}+r^{\frac{1}{\alpha}}B_{\ell^\alpha}\big).$$

Therefore,

$$\Phi(x+\sqrt{r}y+rz) \in A+K\big((C\log n)^{\frac{1}{\alpha}-1}(\sqrt{r}B_{\ell^2}+rB_{\ell^1})+r^{\frac{1}{\gamma}}B_{\ell^\gamma}+r^{\frac{1}{\alpha}}B_{\ell^\alpha}+r^{\frac{1}{2\alpha}}B_{\ell^{2\alpha}}\big).$$

We now simplify the enlargement on the right-hand side. Observe that for any $0<a\le b\le c$,

$$r^{1/b}B_{\ell^b} \subset r^{1/a}B_{\ell^a}+r^{1/c}B_{\ell^c}.$$

Indeed, if $x \in r^{1/b}B_{\ell^b}$, then

$$\sum_{|x_i|\ge 1}|x_i|^a \le \sum_{|x_i|\ge 1}|x_i|^b \le r,$$

and

$$\sum_{|x_i|\le 1}|x_i|^c \le \sum_{|x_i|\le 1}|x_i|^b \le r.$$

Thus, $x = x\mathbb{1}_{x\ge 1}+x\mathbb{1}_{x<1}$, with $x\mathbb{1}_{|x|\ge 1} \in r^{1/a}B_{\ell^a}$ and $x\mathbb{1}_{|x|<1} \in r^{1/c}B_{\ell^c}$. Therefore, as $\alpha \le 2\alpha \le 2$, $\alpha \le \gamma \le 2\alpha$, and $C\log n \ge 1$,

$$\Phi(x+\sqrt{r}y+rz) \in A+K\big((C\log n)^{\frac{1}{\alpha}-1}(\sqrt{r}B_{\ell^2}+rB_{\ell^1})+r^{\frac{1}{\alpha}}B_{\ell^\alpha}\big).$$

Thus,

$$\Phi\big(A+\sqrt{r}B_{\ell^2}+rB_{\ell^1}\big) \subset A+K\big((C\log n)^{\frac{1}{\alpha}-1}(\sqrt{r}B_{\ell^2}+rB_{\ell^1})+r^{\frac{1}{\alpha}}B_{\ell^\alpha}\big). \tag{2.10}$$

Applying the deviation inequality (2.2) of μ_1^n, we get

$$\mu_1^n\big(x \notin A'+\sqrt{r}B_{\ell^2}+rB_{\ell^1}\big) \le \frac{e^{-cr}}{\mu_1^n(A')}.$$

But, since

$$\int ||x||_\infty d\mu_1^n(x) \le c_0 \log n,$$

for some numerical constant $c_0 > 0$, we have by Markov's inequality

$$\mu_1^n(A') = \mu_1^n(A) - \mu_1^n(B^c) \geq \mu_1^n(A) - \frac{c_0}{C}.$$

Thus,

$$\mu_1^n\big(x \notin A' + \sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) \leq \frac{e^{-cr}}{\mu_1^n(A) - c_0/C}.$$

But, as $\mu_\alpha^n = \mu_1^n \circ \Phi^{-1}$, and Φ is a bijection,

$$\begin{aligned}\mu_1^n\big(x \notin A' + \sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) &= \mu_\alpha^n\big(\Phi(\mathbb{R}_+^n \setminus (A' + \sqrt{r}B_{\ell^2} + rB_{\ell^1})\big) \\ &= \mu_\alpha^n\big(\mathbb{R}_+^n \setminus \Phi(A' + \sqrt{r}B_{\ell^2} + rB_{\ell^1})\big).\end{aligned}$$

Using (2.10), we deduce

$$\mu_\alpha^n\big(x \notin A + K\big((C\log n)^{\frac{1}{\alpha}-1}(\sqrt{r}B_{\ell^2} + rB_{\ell^1}) + r^{\frac{1}{\alpha}}B_{\ell^\alpha}\big)\big) \leq \frac{e^{-cr}}{\mu_1^n(A) - c_0/C}.$$

Adjusting the constant c we get the claim. □

The monotone rearrangement ψ of ν_1 onto ν_α, is linked to φ by the relation

$$\forall x \in \mathbb{R},\ \ \psi(x) = \mathrm{sg}(x)\varphi(|x|),$$

where $\mathrm{sg}(x)$ denotes the sign of x. Thus, the monotone rearrangement of ν_1 onto ν_α satisfies the same estimate of Lemma 2.2.4 as φ. Therefore, the same arguments as for the proof of Proposition 2.2.2 can be carried out, and yield the same deviation inequality as for μ_α.

2.2.5 Proposition. *Let $n \in \mathbb{N}$. There is a constant $c > 0$ depending on α, such that for any $r > 0$, A Borel subset of $\mathbb{R}^n$, and $C > 0$ such that $\nu_\alpha^n(A) > 1/C$,*

$$\nu_\alpha^n\Big(x \notin A + C(\log n)^{\frac{1}{\alpha}-1}\big(\sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) + r^{\frac{1}{\alpha}}B_{\ell^\alpha}\Big) \leq \frac{e^{-cr}}{\nu_\alpha^n(A) - 1/C}.$$

In view of this deviation inequality for ν_α, we define a concentration property $\mathcal{C}_\alpha$ when $\alpha < 1$, by saying that a random Hermitian matrix X satisfies the *concentration property* $\mathcal{C}_\alpha$ if there is some $\kappa > 0$, such that for any $r > 0$ and A Borel subset of $\mathcal{H}_n^{(\beta)}$ with $\mathbb{P}(X \in A) \geq 1/2$,

$$\mathbb{P}\Big(X \notin A + \kappa(\log n)^{\frac{1}{\alpha}-1}\big(\sqrt{r}B_{\ell^2} + rB_{\ell^1}\big) + \kappa r^{\frac{1}{\alpha}}B_{\ell^\alpha}\Big) \leq 4e^{-r}. \tag{2.11}$$

As for the case where $\alpha \in [1,2]$, the above concentration property can be translated to a deviation inequality for Lipschitz or Hölder functions as stated in the following lemma.

2.2.6 Lemma. *Let $\alpha \in (0,1)$. Assume X satisfies the concentration property $\mathcal{C}_\alpha$ for some $\kappa > 0$. Let $f : \mathcal{H}_n^{(\beta)} \to \mathbb{R}$ be a function respectively L_1-Lipschitz and L_2-Lipschitz with respect to $\| \ \|_{\ell^1}$, and $\| \ \|_{\ell^2}$. There is a constant $c > 0$ depending* *on α, such that if f is moreover L_α-Lipschitz with respect to $|| \, ||_{\ell^\alpha}^\alpha$, then for any $t > 0$,*

$$\mathbb{P}\big(f(X) > m_f + t\big) \leq 4 \exp\Big(- c \min\Big(\frac{t^2}{\kappa^2 (\log n)^{2(\frac{1}{\alpha}-1)} L_2}, \frac{t}{\kappa (\log n)^{\frac{1}{\alpha}-1} L_1 + \kappa L_\alpha} \Big)\Big),$$

whereas if

$$\forall A, B \in \mathcal{H}_n^{(\beta)}, \; f(A) - f(B) \leq L'_\alpha ||A - B||_{\ell^\alpha},$$

for some $L'_\alpha > 0$, then for any $t > 0$,

$$\mathbb{P}\big(f(X) > m_f + t\big) \leq 4 \exp\Big(- c \min\Big(\frac{t^2}{\kappa^2 (\log n)^{2(\frac{1}{\alpha}-1)} L_2}, \frac{t}{\kappa (\log n)^{\frac{1}{\alpha}-1} L_1}, \frac{t^\alpha}{\kappa^\alpha L'^\alpha_\alpha} \Big)\Big),$$

where m_f is the median of $f(X)$.

2.3 Concentration inequality for the spectral measure

We denote by d the following distance on the set of probability measures on $\mathbb{R}$, denoted by $\mathcal{P}(\mathbb{R})$,

$$\forall \mu, \nu \in \mathcal{P}(\mathbb{R}), \; d(\mu, \nu) = \sup_{z \in \mathcal{K}} |g_\mu(z) - g_\nu(z)|, \tag{2.12}$$

where $\mathcal{K}$ is a compact subset of $\{z \in \mathbb{C} : \Im z \geq 1\}$ with an accumulation point, such that $\operatorname{diam}(K) \leq 1$, and with g_μ the Stieltjes transform of μ. Let $\mathcal{W}_p$ denote the L^p-Wasserstein distance, for $p \geq 1$, by

$$\mathcal{W}_p(\mu, \nu) = \Big(\inf_\pi \int |x - y|^p d\pi(x, y) \Big)^{1/p},$$

and for $0 < p < 1$,

$$\mathcal{W}_p(\mu, \nu) = \inf_\pi \int |x - y|^p d\pi(x, y),$$

where the infimum is taken on all coupling π between μ and ν. As a consequence of the Kantorovitch-Rubinstein duality (see [99, Particular case 5.16]), we have

$$d \leq \mathcal{W}_1.$$

Besides, Jensen's inequality yields for any $p \geq 1$,

$$\mathcal{W}_1 \leq \mathcal{W}_p,$$

Therefore,

$$d \leq \mathcal{W}_1 \leq \mathcal{W}_p. \tag{2.13}$$

In this section, we aim at proving the following concentration inequality for the spectral measure of Wigner matrices satisfying the concentration property $\mathcal{C}_\alpha$, with respect to the distance d defined above.

2.3.1 Proposition. *Let $\alpha \in (0,2]$. Let X be a Wigner matrix satisfying $\mathcal{C}_\alpha$ with some $\kappa > 0$. There exists a constant $c_\alpha > 0$, depending on α, such that for any $t > 0$,*

$$\mathbb{P}\Big(d\big(\mu_{X/\sqrt{n}}, \mathbb{E}\mu_{X/\sqrt{n}}\big) > t + \delta_n\Big) \leq \frac{32}{t^2}\exp\big(-c_\alpha h_\alpha(t)\big),$$

where $\delta_n = O\big(\kappa n^{-1}(\log n)^{(1/\alpha-1)_+}\big)$, and where for $\alpha \in [1,2)$,

$$h_\alpha(t) = \min\Big(\frac{n^2t^2}{\kappa^2}, \frac{n^{1+\frac{\alpha}{2}}t^\alpha}{\kappa^\alpha}\Big),$$

whereas for $\alpha \in (0,1)$

$$h_\alpha(t) = \min\Big(\frac{n^2t^2}{\kappa^2(\log n)^{2(\frac{1}{\alpha}-1)}}, \frac{n^{1+\frac{\alpha}{2}}t}{\kappa}\Big).$$

This concentration inequality enables us to retrieve the speed of large deviation of the spectral measure of Wigner matrices without Gaussian tail. Indeed, by [29], we know that if the tail distributions of the entries behave as $e^{-c|x|^\alpha}$ for some $c > 0$ and $\alpha < 2$, then a LDP holds with speed $n^{1+\alpha/2}$.

In view of Lemmas 2.2.1 and 2.2.6, Proposition 2.3.1 requires to compute the Lipschitz constants of the Stieltjes transform of the spectral measure of Hermitian matrices, with respect to $|| \, ||_{\ell^p}$ when $p \in [1,2]$, and $|| \, ||_{\ell^p}^p$ when $p \in (0,1)$.

2.4 Two lemmas on spectral variation of Hermitian matrices

To compute such Lipschitz constants, we need some inequalities about spectral variation of Hermitian matrices. When $p \in [1,2]$, we have the following inequality which is a direct consequence of Lidskii's theorem (see [22, Corollary III 4.2]).

2.4.1 Lemma. *Let $p \in [1,2]$, and $A, B \in \mathcal{H}_n^{(\beta)}$.*

$$\mathcal{W}_p(\mu_A, \mu_B) \leq \frac{1}{n^{1/p}}||A - B||_{\ell^p}. \tag{2.14}$$

As a consequence,

$$d(\mu_A, \mu_B) \leq \frac{1}{n^{1/p}}||A - B||_{\ell^p}.$$

Proof. By Lidskii's theorem (see [22, Corollary III 4.2]), we have

$$\lambda^\downarrow(A) - \lambda^\downarrow(B) \prec \lambda^\downarrow(A - B),$$

where $\lambda^\downarrow(A)$ denotes the vector of eigenvalues of A in decreasing order, and $\prec$ denotes the Löwner order between vectors of $\mathbb{R}^n$ (see [22, Corollary III 4.2]). Thus, by [22, Theorem II.3.1] we get, since $x \mapsto |x|^p$ is convex as $p \geq 1$,

$$\mathrm{tr}|\lambda^\downarrow(A) - \lambda^\downarrow(B)|^p \leq \mathrm{tr}|\lambda^\downarrow(A - B)|^p.$$

Using the decreasing coupling between the spectra of A and B, we get

$$\mathcal{W}_p(\mu_A, \mu_B) \leq \frac{1}{n^{1/p}}||A - B||_p, \tag{2.15}$$

where $|| \ ||_p$ denotes the p-Schatten norm on $\mathcal{H}_n^{(\beta)}$, which is defined by,

$$\forall A \in \mathcal{H}_n^{(\beta)}, \ ||A||_p = \left(\mathrm{tr}|A|^p\right)^{1/p}, \tag{2.16}$$

if $p < +\infty$, and by setting $|| \ ||_\infty$ to be the spectral radius. But as $p \leq 2$, we have by [104, Theorem 3.32],

$$||A - B||_p \leq ||A - B||_{\ell^p}, \tag{2.17}$$

which ends the proof of the lemma. □

2.4.2 Remark. When $p > 2$, inequality (2.14) is no longer true, since for $B = 0$ it amounts to (2.17), which is false when $p > 2$, by taking $A = uu^*$, where u is the constant vector. But we still have

$$\mathcal{W}_p(\mu_A, \mu_B) \leq \frac{1}{n^{1/p}}||\lambda(A - B)||_{\ell^p},$$

which is not workable for concentration purposes in the setting of Wigner matrices. This said, as already remarked before, when $\alpha \geq 2$, ν_α^n satisfies the log-Sobolev inequality, thus one has a sub-Gaussian concentration inequality for the resolvent of Wigner matrices with such laws.

When $p \leq 1$, one may hope for the inequality

$$\mathcal{W}_p(\mu_A, \mu_B) \leq \frac{1}{n}||A - B||_{\ell^p}^p, \tag{2.18}$$

to hold. But taking formally $p \to 0$, would yield

$$|\lambda(A)\Delta\lambda(B)| \leq |(i,j) : A_{i,j} \neq B_{i,j}|,$$

where $\lambda(A), \lambda(B)$ denote the set of eigenvalues of A and B. But changing 1 entry to a matrix can change the whole spectrum. Indeed, if X has only simple eigenvalues λ_i associated to unit eigenvectors v_i, then if one takes a "delocalized" perturbation uu^* with $\langle v_i, u\rangle \neq 0$, then one can see that the equation,

$$1 - \langle u, (x - X)^{-1}u\rangle = 1 - \sum_{i=1}^{n} \frac{|\langle u, e_i\rangle|^2}{x - \lambda_i} = 0,$$

admit n solutions. We deduce from (1.7) that the spectrum of $X + uu^*$ is disjoint from the one of X. Thus, if u is a coordinate vector, this argument gives no hope to (2.18) to be true.

The moral of remark 2.4.2 is that one cannot have (2.18) with a constant 1 on the right-hand side. As the cost function $| \ |^p$ behaves quite badly when $p < 1$ as it is not convex (see [42] for the assignment problem), in particular, the optimal transport map is not necessarily the monotone rearrangement contrary the case $p \geq 1$ (see [98, Theorem 2.18, remark 2.19 (ii)]), we will note investigate further the question of having a spectral variation inequality involving the L^p-Wasserstein distance. We prefer to deal with another distance on $\mathcal{P}(\mathbb{R})$ which induces the same topology as $\mathcal{W}_p$ and dominates d. This distance is chosen so that, applied to empirical spectral measures, it will be controlled by $|| \, ||_{\ell^p}^p$ in the case where $p \in (0,1)$.

To this end, let $p \in (0,1)$ and define the subset $\mathcal{P}_p(\mathbb{R})$ of probability measures on $\mathbb{R}$ with finite p^{th} moments. For any $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$, we set

$$d_p(\mu,\nu) = \sup_{t\in\mathbb{R}} \Big| \int (t-x)_+^p d\mu(x) - \int (t-x)_+^p d\nu(x) \Big|. \tag{2.19}$$

Taking formally p to 0, we retrieve the Kolmogorov-Smirnov distance d_{KS}. Recall that by integrating by parts, we can write

$$d_{KS}(\mu,\nu) = \sup \Big\{ \Big| \int f d\mu - \int f d\nu \Big| : f \in \mathrm{NBV}, \ ||f||_{BV} \leq 1 \Big\},$$

where NBV denotes the set of normalized functions with bounded variation, that is, functions which are the integrals of finite signed measures, and

$$||f||_{BV} = |\sigma|(\mathbb{R}),$$

where f is the distribution function of the finite signed measure σ, and $|\sigma|$ is its total variation.

We can actually have a similar formulation for d_p, by introducing the fractional integral operators of order $p+1$ on the set $\mathcal{M}_s^p$ of finite signed probability measures σ such that $|\sigma|$ has a finite p^{th} moment. Following [86, Chapter 2 §5.1], we define for $\mu \in \mathcal{M}_s^p$,

$$\begin{aligned} \forall t \in \mathbb{R}, \ (\mathcal{I}_+^{p+1}\mu)(t) &= \frac{1}{\Gamma(p+1)} \int_{-\infty}^{t} (t-x)^p d\mu(x), \\ (\mathcal{I}_-^{p+1}\mu)(t) &= \frac{1}{\Gamma(p+1)} \int_{t}^{+\infty} (x-t)^p d\mu(x). \end{aligned}$$

This definition interpolates for non-integer order the usual iterated integral. With this notation, we have the following integration by parts formula, for μ and ν finite signed measure with finite p^{th}-moment,

$$\int (\mathcal{I}_+^{p+1}\mu)(t) d\nu(t) = \int (\mathcal{I}_-^{p+1}\nu)(x) d\mu(x). \tag{2.20}$$

Thus,

$$\begin{aligned} d_p(\mu,\nu) &= \Gamma(p+1) \sup_{t\in\mathbb{R}} \big| (\mathcal{I}_+^{p+1}\mu)(t) - (\mathcal{I}_+^{p+1}\mu)(t) \big| \\ &= \Gamma(p+1) \sup_{\sigma} \Big| \int (\mathcal{I}_-^{p+1}\sigma) d\mu - \int (\mathcal{I}_-^{p+1}\sigma) d\nu \Big|, \end{aligned} \tag{2.21}$$

where the supremum is taken on all $\sigma \in \mathcal{M}_s^p$, such that $|\sigma|(\mathbb{R}) \leq 1$. The inequality $d_p \geq$ (2.21) is the consequence of the integration by parts formula (2.20), whereas the equality is given by taking $\sigma = \delta_t$, for $t \in \mathbb{R}$. We investigate now the link between the distances d, defined in (2.12), $\mathcal{W}_p$ and d_p for $p \in (0,1)$.

2.4.3 Proposition. *Let $p \in (0,1)$. Then, d_p, defined in (2.19), is a distance on $\mathcal{P}_p(\mathbb{R})$, and metrizes the weak topology. More precisely, there is a constant $C_p > 0$ such that*

$$d(\mu,\nu) \leq C_p d_p(\mu,\nu), \tag{2.22}$$

for all $\mu,\nu \in \mathcal{P}_p(\mathbb{R})$. One can choose

$$C_p = \sqrt{\pi}(p+1)\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(1+\frac{p}{2}\right)}. \tag{2.23}$$

Furthermore,

$$d_p \leq \mathcal{W}_p. \tag{2.24}$$

Proof. In view of the formulation of d_p as (2.21), the stake behind (2.22) is to represent the function $t \mapsto (z-t)^{-1}$ as the fractional integral of order $p+1$ of some function. The constant C_p will arise as the L^1 norm of this function as $\Im z \geq 1$, over $\Gamma(p+1)$.

The fractional integral of order $p+1$ of the function $t \mapsto (z-t)^{-1}$ is given in [86, Chapter 2 (5.25)], which we state in the next lemma.

2.4.4 Lemma. *Let $p \in (0,1)$. For any $z \in \mathbb{C}$, $\Im z > 0$, we have*

$$\forall x \in \mathbb{R},\ \frac{1}{z-x} = \mathcal{I}_-^{p+1}(\varphi)(x),$$

with

$$\forall t \in \mathbb{R},\ \varphi(t) = e^{i\pi(p+1)}\Gamma(p+2)\frac{1}{(z-t)^{p+2}}, \tag{2.25}$$

where ζ^p is the principal branch of the α^{th}-root on $\mathbb{C}\setminus\mathbb{R}_-$.

Let $\Im z \geq 1$ and φ as in (2.25). We have

$$\frac{1}{\Gamma(p+1)}||\varphi||_1 \leq (p+1)\int \frac{dt}{(1+t^2)^{1+p/2}} := C_p,$$

where we used $\Gamma(p+2) = (p+1)\Gamma(p+1)$. Therefore,

$$d \leq C_p d_p.$$

But, one can recognize an Euler integral of the first kind in the definition of C_p, by making successively the changes of variables $t = \tan u$, and $v = (\cos u)^2$, which yields,

$$C_p = (p+1)\int_0^1 v^{\frac{p-1}{2}}(1-v)^{-\frac{1}{2}}dv.$$

Therefore by [5, (2.13)], we deduce the value for C_p claimed in (2.23). Inequality (2.24) is the consequence of the sub-additivity of the function $x \mapsto x^p$ on $\mathbb{R}^+$. More precisely, for any $x,y,t \in \mathbb{R}$,

$$(t-x)_+^p - (t-y)_+^p \leq |x-y|^p.$$

Integrating the above inequality under a coupling P of two probability measures with finite p^{th}-moment yields the claim.

From (2.22), we deduce that the topology induced by d_p on $\mathcal{P}_p(\mathbb{R})$ is finer than the weak topology, and by (2.24) that it is coarser than the one induced by $\mathcal{W}_p$. But $\mathcal{W}_p$ induces the weak topology on $\mathcal{P}_p(\mathbb{R})$ by [99, Theorem 6.9] (as $|\ |^p$ is a metric on $\mathbb{R}$ for $p \leq 1$), therefore d_p induces the weak topology on this set. □

We finally prove that the distance d_p we introduced, when applied to spectral measures of Hermitian matrices, is dominated by $||\ ||_{\ell^p}^p$ for $p \in (0,1)$.

2.4.5 Lemma. *Let $p \in (0,1)$. Let $A, B \in \mathcal{H}_n^{(\beta)}$. For any $t \in \mathbb{R}$,*

$$\Big|\sum_{i=1}^n (t-\lambda_i(A))_+^p - \sum_{i=1}^n (t-\lambda_i(B))_+^p\Big| \leq \sum_{i=1}^n |\lambda_i(A-B)|^p. \tag{2.26}$$

Therefore,

$$d_p(\mu_A, \mu_B) \leq \frac{1}{n}\sum_{i=1}^n |\lambda_i(A-B)|^p \leq \frac{1}{n}\sum_{i,j} |A_{i,j} - B_{i,j}|^p, \tag{2.27}$$

where d_p is defined in (2.19). In particular,

$$d(\mu_A, \mu_B) \leq \frac{C_p}{n}||A-B||_{\ell^p}^p, \tag{2.28}$$

where C_p is as in (2.23).

Proof. As $\alpha \leq 2$, the second inequality of (2.27) is due to [104, Theorem 3.32]. To prove the first inequality (2.12), we begin by recalling an inequality due to Rotfel'd originally, and then to Thompson [96] (for an extension and a simpler proof). Let $F : \mathbb{R}_+^{2n} \to \mathbb{R}$ be a concave symmetric function. Then for any $A, B \in \mathcal{H}_n^{(\beta)}$ positive semi-definite,

$$F(\lambda(A+B), 0) \leq F(\lambda(A), \lambda(B)),$$

where $\lambda(C)$ denotes the vector of eigenvalues of a Hermitian matrix C. Note that since F is symmetric, there is no ambiguity in the writing. Let $t \in \mathbb{R}$. We have,

$$t - A - B \leq (t-A)_+ + |B|.$$

In particular, if we denote $\lambda_1(C) \geq \lambda_2(C) \geq ... \geq \lambda_n(C)$ the eigenvalues of some Hermitian matrix C, then by Weyl's inequality (1.6), for any $i \in \{1, ..., n\}$,

$$\lambda_i(t-A-B) \leq \lambda_i\big((t-A)_+ + |B|\big).$$

Therefore,

$$\lambda_i(t-A-B)_+ \leq \lambda_i\big((t-A)_+ + |B|\big).$$

Define

$$\forall x \in \mathbb{R}_+^{2n},\ F(x) = \sum_{i=1}^{2n} x_i^\alpha.$$

Since A, B are Hermitian,

$$\lambda(t-A-B)_+ = (t - \lambda(A+B))_+.$$

As F is non-decreasing coordinate-wise,

$$F\big((t-\lambda(A+B))_+,0\big) \leq F\big(\lambda((t-A)_+ + |B|),0\big).$$

Rotfel'd inequality gives

$$F\big(\lambda((t-A)_+ + |B|),0\big) \leq F\big((t-\lambda(A))_+,|\lambda(B)|\big).$$

Thus,

$$\sum_{i=1}^n \big(t-\lambda_i(A+B)\big)_+^\alpha \leq \sum_{i=1}^n \big(t-\lambda_i(A)\big)_+^\alpha + \sum_{i=1}^n |\lambda_i(B)|^\alpha.$$

Applying this inequality with $A+B$, $-B$ instead of A and B, we get the claim. □

With the Lemmas 2.4.5 and 2.4.1, we can now give a proof of Proposition 2.3.1.

Proof of Proposition 2.3.1. Fix some $z \in \mathcal{K}$. Let f denote the function on $\mathcal{H}_n^{(\beta)}$ defined by,

$$\forall Y \in \mathcal{H}_n^{(\beta)},\ f_z(Y) = g_{\mu_{Y/\sqrt{n}}}(z).$$

As by (2.13), $d \leq \mathcal{W}_p$, for $p \geq 1$, we deduce from Lemma 2.4.1 that f_z is $n^{-\frac{1}{p}-\frac{1}{2}}$-Lipschitz with respect to $|| \, ||_{\ell^p}$, when $p \in [1,2]$. If $p<1$, Lemma 2.4.5 tells us that f_z is $C_p n^{-1-\frac{p}{2}}$-Lipschitz with respect to $|| \, ||_{\ell^p}^p$.

Let m_z be the median of $f_z(X)$. Therefore, by Lemma 2.2.1 or Lemma 2.2.6, there is a constant $c_\alpha > 0$ such that for any $t>0$,

$$\mathbb{P}(|f_z - m_z| > t) \leq 8\exp\big(-c_\alpha h_\alpha(t)\big), \tag{2.29}$$

with

$$h_\alpha(t) = \min\Big(\frac{t^2n^2}{\kappa^2}, \frac{t^\alpha n^{1+\frac{\alpha}{2}}}{\kappa^\alpha}\Big),$$

when $\alpha \in [1,2]$, and

$$h_\alpha(t) = \min\Big(\frac{t^2n^2}{\kappa^2(\log n)^{2(\frac{1}{\alpha}-1)}}, \frac{t}{\kappa(\log n)^{\frac{1}{\alpha}-1}n^{-\frac{3}{2}} + \kappa n^{-1-\frac{\alpha}{2}}}\Big),$$

for $\alpha \in (0,1)$. In the case $\alpha<1$, note that one can find a constant $c>0$ such that

$$h_\alpha(t) \geq c\min\Big(\frac{t^2n^2}{\kappa^2(\log n)^{2(\frac{1}{\alpha}-1)}}, \frac{tn^{1+\frac{\alpha}{2}}}{\kappa}\Big).$$

Integrating the inequality (2.29), we get

$$|\mathbb{E}f_z(X) - m_z| \leq \varepsilon_n$$

with $\varepsilon_n = O(\kappa n^{-1})$, if $\alpha \in [1,2]$ and $\varepsilon_n = O(\kappa n^{-1}(\log n)^{\frac{1}{\alpha}-1})$, if $\alpha < 1$, uniformly in $z \in \mathbb{C}$, $\Im z \geq 1$. With this notation, we get for any $t>0$,

$$\mathbb{P}(|f_z - \mathbb{E}f_z| > t + \varepsilon_n) \leq 8\exp\big(-c_\alpha h_\alpha(t)\big).$$

Let $\mathcal{N}_t$ be a t-net of $\mathcal{K}$. As $z \mapsto f_z(X)$ is 1-Lipschitz on $\{z \in \mathbb{C} : \Im z \geq 1\}$, we have

$$\mathbb{P}\big(\sup_{z\in\mathcal{K}} |f_z - \mathbb{E}f_z| > 2t + \varepsilon_n\big) \leq 8|\mathcal{N}_t|\exp\big(-c_\alpha h_\alpha(t)\big),$$

As $\mathcal{K}$ is a subset of $\mathbb{C}$ of diameter inferior to 1, we can find a t-net $\mathcal{N}_t$ such that $|\mathcal{N}_t| \leq t^{-2}$. Thus,

$$\mathbb{P}\big(d(\mu_X, \mathbb{E}\mu_X) > 2t + \varepsilon_n\big) \leq \frac{8}{t^2} \exp\big(-c_\alpha h_\alpha(t) n\big),$$

which, adjusting the constant c_α, gives the claim.

□

2.5 Concentration inequality for the largest eigenvalue

We provide now a deviation inequality for the largest eigenvalue of Wigner matrices satisfying the concentration property $\mathcal{C}_\alpha$. We denote by λ_Y the largest eigenvalue of $Y \in \mathcal{H}_n^{(\beta)}$.

2.5.1 Proposition. *Let $\alpha \in (0, 2]$. Let X be a Wigner matrix satisfying $\mathcal{C}_\alpha$ for some $\kappa > 0$. There is a constant $c_\alpha > 0$, such that for any $t > 0$,*

$$\mathbb{P}\left(\left|\lambda_{X/\sqrt{n}} - \mathbb{E}\lambda_{X/\sqrt{n}}\right| > t + \varepsilon_n\right) \leq 8 \exp\big(-c_\alpha h_\alpha(t)\big),$$

where

$$h_\alpha(t) = \min\Big(\frac{nt^2}{\kappa^2}, \frac{n^{\frac{\alpha}{2}} t^\alpha}{\kappa^\alpha}\Big), \tag{2.30}$$

if $\alpha \in [1, 2]$, and

$$h_\alpha(t) = \min\Big(\frac{nt^2}{\kappa^2 (\log n)^{2(\frac{1}{\alpha}-1)}}, \frac{\sqrt{n} t}{\kappa (\log n)^{\frac{1}{\alpha}-1}}, \frac{n^{\frac{\alpha}{2}} t^\alpha}{\kappa^\alpha}\Big), \tag{2.31}$$

if $\alpha \in (0, 1)$, and where $\varepsilon_n = O(\kappa n^{-1/2} (\log n)^{(1/\alpha - 1)_+})$.

Proof. By Weyl's inequality (1.6), the function

$$f : Y \in \mathcal{H}_n^{(\beta)} \mapsto \lambda_{Y/\sqrt{n}}$$

is $n^{-1/2}$-Lipschitz with respect to $|| \, ||_p$, for any $p > 0$. Let m_f denote the median of $f(X)$, and $t > 0$. As $\alpha \leq 2$, $|| \, ||_\alpha \leq || \, ||_{\ell^\alpha}$ (see [104, Theorem 3.32]), f is also 1-Lipschitz with respect to $|| \, ||_{\ell^\alpha}$. By Lemmas 2.2.1 and 2.2.6, in the case $\alpha < 1$, we deduce that for any $t > 0$,

$$\mathbb{P}(|f - m_f| > t) \leq 8 \exp\big(-c_\alpha h_\alpha(t)\big), \tag{2.32}$$

with h_α as in (2.30), (2.31). Integrating the above inequality, we get

$$|\mathbb{E} f(X) - m_f| = O(\kappa n^{-1/2} (\log n)^{(\frac{1}{\alpha}-1)_+}), \tag{2.33}$$

which gives the claim.

□

2.6 Concentration for non-commutative polynomials

We show, in this last subsection, concentration inequalities for non-commutative polynomials of a family of Hermitian matrices satisfying the concentration property $\mathcal{C}_\alpha$, when $\alpha \in [1,2]$, and q deterministic Hermitian matrices. As the trace of polynomials of random matrices is not a Lipschitz function, we need to take some care in applying the concentration property $\mathcal{C}_\alpha$. We mention the work of Meckes-Szarek [79] who gave a concentration inequality for polynomials under a normal concentration assumption of the law of the matrices, using a truncation-optimization approach. We will follow here another road which takes advantage of the "stability" of our functional, which is similar to the concentration inequality for Gaussian chaoses Ledoux and Talagrand gave in [73, Lemma 3.8].

In the following, we denote for $\mathbf{Y} = (Y_1, ..., Y_p)$ and for any $k \in (0, +\infty]$,

$$||\mathbf{Y}||_k = || \oplus_{k=1}^p Y_i||_k,$$

where $||\ ||_k$ denotes the k-Schatten norm on Hermitian matrices defined in (2.16). We will say that a family of random Hermitian matrices $\mathbf{X} = (X_1, ..., X_p)$ satisfies the *concentration property* $\mathcal{C}_\alpha$ for some $\alpha \in [1,2]$, if $Y = \oplus_{i=1}^p X_i$ satisfies $\mathcal{C}_\alpha$ as defined in (2.1).

2.6.1 Proposition. *Let $\alpha \in [1,2]$. Let P be a non commutative polynomial in $p+q$ variables, with total degree d in the first p variables. Let $\boldsymbol{X} = (X_1, ..., X_p)$ be independent Hermitian random matrices satisfying $\mathcal{C}_\alpha$ with some constant $\kappa > 0$, and $\boldsymbol{D} = (D_1, ..., D_q)$ deterministic Hermitian matrices with spectral radii bounded by $\rho \geq 1$. Set $L_2, L_\alpha \geq 1$ such that*

$$\mathbb{P}\big(||\boldsymbol{X}||_{2(d-1)}^{d-1} \leq \kappa^{d-1} L_2,\ ||\boldsymbol{X}||_{\alpha'(d-1)}^{d-1} \leq \kappa^{d-1} L_\alpha\big) \geq \frac{1}{c_0} + \frac{1}{2}, \tag{2.34}$$

with some $c_0 > 0$, and where α' the conjugate exponent of α. There are some constants $c, C > 0$ depending on P, such that for any $t > 0$,

$$\begin{aligned}\mathbb{P}\Big(\big|\mathrm{tr}[P(\boldsymbol{X},\boldsymbol{D})] - \mathbb{E}\mathrm{tr}[P(\boldsymbol{X},\boldsymbol{D})]\big| > (t+\varepsilon_n)\rho^q\kappa^d\Big)& \\ \leq c_0 C \exp\Big(-c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}\Big)\Big)&,\end{aligned}$$

with $\varepsilon_n = O(\kappa^d \rho^q (L_2 + L_\alpha))$.

2.6.2 Remark. As we are unable to understand the "stability" of the trace of a polynomial with respect to the distance $||\ ||_{\ell^\alpha}^\alpha$ for $\alpha < 1$, that is, to prove the equivalent of Lemma 2.6.4 below when $\alpha < 1$, we cannot provide a deviation inequality in the case $\alpha < 1$.

Applying Proposition 2.6.1 to Wigner matrices, we get the following inequality.

2.6.3 Proposition. *Let $\alpha \in [1,2]$. Let P be a non commutative polynomial in $p+q$ variables, with total degree d in the first p variables. Let $\boldsymbol{X} = (X_1, ..., X_p)$ be*

Wigner matrices satisfying $\mathcal{C}_\alpha$, and $\boldsymbol{D} = (D_1, ..., D_q)$ be deterministic Hermitian matrices with spectral radii bounded by $\rho \geq 1$. We assume that

$$||\mathbb{E}\boldsymbol{X}||_{\alpha'(d-1)}^{d-1} \leq r\kappa^{d-1}n^{\theta},\ ||\mathbb{E}\boldsymbol{X}||_{2(d-1)}^{d-1} \leq r\kappa^{d-1}n^{d/2}, \tag{2.35}$$

where $r \geq 1$ and $\theta = \frac{d}{2} - \frac{1}{2} + \frac{1}{k}$. There are some constant $c, C > 0$ depending on P such that for any $t > 0$,

$$\mathbb{P}\Big(\big|\tau_n[P(\boldsymbol{X}/\sqrt{n}, \boldsymbol{D})] - \mathbb{E}\tau_n[P(\boldsymbol{X}/\sqrt{n}, \boldsymbol{D})]\big| > (t + \eta_n)\kappa\rho^q\Big)$$
$$\leq C\exp\Big(- c\min\Big(n^2t^2, n^{1+\frac{\alpha}{2}}t^{\alpha}, n^{\alpha(\frac{1}{2}+\frac{1}{d})}t^{\frac{\alpha}{d}}\Big)\Big),$$

with τ_n standing for $\frac{1}{n}\mathrm{tr}$ and $\eta_n = O(\kappa^d\rho^q rn^{-\frac{1}{2}-\frac{1}{\alpha}})$.

One can interpret this three-levels deviation inequality as the following. The Gaussian part in n^2t^2 allows to capture the Gaussian fluctuations of the traces of polynomials in independent Wigner matrices,

$$n\big(\tau_n[P(\mathbf{X}/\sqrt{n}, \mathbf{D})] - \mathbb{E}\tau_n[P(\mathbf{X}/\sqrt{n}, \mathbf{D})]\big),$$

known in particular for the GUE case by [91, Theorem 13], and for $q = 0$ and $p = 1$ for Wigner matrices by [3, Theorem 2.1.31]. The term is $n^{1+\alpha/2}t^{\alpha}$ reveals the participation of the spectral measure in the deviations of the trace. Indeed, as we saw in Proposition 2.3.1, $n^{1+\alpha/2}$ is the speed of deviations of the empirical spectral measure of Wigner matrices satisfying $\mathcal{C}_\alpha$. Some deviations of the trace of certain polynomials are due to the empirical spectral measure, as one can see in the simple example where $p = 1$, $q = 0$, and $P = X^{2k}$, we have

$$\mathbb{P}\big(\mu_{X/\sqrt{n}}(x^{2k}) < \mathbb{E}\mu_{X/\sqrt{n}}(x^{2k}) + t\big) \leq \mathbb{P}\big(d(\mu_{X/\sqrt{n}}, \mathbb{E}\mu_{X/\sqrt{n}}) > h(t)\big),$$

for some function h such that $h(t) \to 0$ as $t \to 0$, and which one can make depend only on $\mathbb{E}X_{1,1}^2$ and $\mathbb{E}X_{1,2}^2$. This is a consequence of the lower semi-continuity of the map $\mu \mapsto \mu(x^{2k})$, and the fact that $\mathbb{E}\mu_{X/\sqrt{n}}$ is in the compact subset of measures with second moment bounded by $\max(2\mathbb{E}|X_{1,2}|^2, \mathbb{E}X_{1,1}^2)$.

Finally, the third level in $n^{\alpha(\frac{1}{2}+\frac{1}{d})}t^{\alpha/d}$ captures the large deviations behavior of the trace of $P(\mathbf{X}/\sqrt{n}, \mathbf{D})$, as we will see in chapter 4.

Before going into the proof of Proposition 2.6.1 we show how to deduce Proposition 2.6.3 from Proposition 2.6.1.

Proof of Proposition 2.6.3. Let $k \in \{2, \alpha'\}$. If $k < +\infty$, we know from [3, Lemma 2.1.6] that, for any $i \in \{1, ..., p\}$,

$$\mathbb{E}\mathrm{tr}|X_i - \mathbb{E}X_i|^{k(d-1)} \leq c_1\kappa^{k(d-1)}n^{1+\frac{k(d-1)}{2}},$$

where $c_1 \geq 1$ is some constant depending on α and d. Thus, by Jensen's inequality,

$$\mathbb{E}||\mathbf{X} - \mathbb{E}\mathbf{X}||_{k(d-1)}^{d-1} \leq \Big(\sum_{i=1}^{p}\mathbb{E}\mathrm{tr}|X_i|^{k(d-1)}\Big)^{1/k} \leq (pc_1)^{1/k}\kappa^{d-1}n^{\theta(k)}, \tag{2.36}$$

with $\theta(k) = \frac{d}{2} - \frac{1}{2} + \frac{1}{k}$. If $k = +\infty$, we know by [3, Theorem 2.1.22] or [10, Theorem 5.1] that there is some constant $c_2 > 0$, such that $\mathbb{E}||\mathbf{X} - \mathbb{E}\mathbf{X}||_\infty \leq c_2 \kappa n^{1/2}$, so that (2.36) holds also for $k = +\infty$, with c_1 big enough, and with the convention $\theta(+\infty) = (d-1)/2$. Thus, using the assumption of (2.35), we get by the triangular inequality and convexity,

$$\mathbb{E}||\mathbf{X}||_{k(d-1)}^{d-1} \leq 2^{d-2}\kappa^{d-1}\big((pc_1)^{1/k} + r\big)n^{\theta(k)},$$

As $r \geq 1$, we get

$$\mathbb{E}||\mathbf{X}/\sqrt{n}||_{k(d-1)}^{d-1} \leq c'\Big(\frac{\kappa}{\sqrt{n}}\Big)^{d-1} r n^{\theta(k)},$$

where $c' > 0$ is some numerical constant. By Markov's inequality, we see that the assumption of Proposition 2.6.1 holds with $L_k = 8c'rn^{\theta(k)}$ for $k \in \{2, \alpha'\}$, and $c_0 = 1/4$, as $\mathbf{X}/\sqrt{n}$ has concentration $\mathcal{C}_\alpha$ with constant $\kappa/\sqrt{n}$. We deduce that for any $s > 0$,

$$\mathbb{P}\Big(\big|\mathrm{tr}P(\mathbf{X}/\sqrt{n}, \mathbf{D}) - \mathbb{E}\mathrm{tr}P(\mathbf{X}/\sqrt{n}, \mathbf{D})\big| > (s + \varepsilon_n)\rho^q\kappa^d n^{-\frac{d}{2}}\Big)$$
$$\leq 4C\exp\Big(-c\min\Big(\frac{s^2}{r^2 n^{2\theta(2)}}, \frac{s^\alpha}{r^\alpha n^{\alpha\theta(\alpha)}}, s^{\frac{\alpha}{d}}\Big)\Big),$$

with $\varepsilon_n = O(\kappa^d\rho^q r n^{\theta(\alpha)})$, and where $c, C > 0$ are constants depending on P. Taking $s = n^{1+\frac{d}{2}}t$, we get

$$\mathbb{P}\Big(\big|\tau_n[P(\mathbf{X}/\sqrt{n}, \mathbf{D})] - \mathbb{E}\tau_n[P(\mathbf{X}/\sqrt{n}, \mathbf{D})]\big| > (t + \eta_n)\rho^q\kappa^d\Big)$$
$$\leq 4C\exp\Big(-c\min\Big(\frac{t^2n^2}{r^2}, \frac{t^\alpha n^{1+\frac{\alpha}{2}}}{r^\alpha}, t^{\frac{\alpha}{d}} n^{\alpha(\frac{1}{d}+\frac{1}{2})}\Big)\Big),$$

with $\eta_n = O(\kappa^d\rho^q r n^{-\frac{1}{2}-\frac{1}{\alpha}})$.

□

Proof of Proposition 2.6.1. We will start with a monomial $Q \in \mathbb{C}\langle\mathbf{X}, \mathbf{D}\rangle$ of total degree d in $\mathbf{X}$. We define the function f by,

$$\forall \mathbf{Y} \in (\mathcal{H}_n^{(\beta)})^p, \ f(\mathbf{Y}) = \mathrm{tr}Q(\mathbf{Y}, \mathbf{D}).$$

In view of the concentration property $\mathcal{C}_\alpha$, we need to understand how the map f gets deformed under enlargements with respect to $||\ ||_2$ and $||\ ||_\alpha$. This is given by the following inequality.

2.6.4 Lemma. *Let $\alpha \in [1,2]$. There is a constant $C_d > 0$, such that for any $\boldsymbol{Y}, \boldsymbol{H} \in (\mathcal{H}_n^{(\beta)})^p$,*

$$|f(\boldsymbol{Y} + \boldsymbol{H}) - f(\boldsymbol{Y})| \leq C_d\rho^q\big(||\boldsymbol{Y}||_{\alpha'(d-1)}^{d-1} + ||\boldsymbol{H}||_\alpha^{d-1}\big)||\boldsymbol{H}||_\alpha,$$

where α' is the conjugate exponent of α, and ρ is a bound on the spectral radii of D.

Proof. By the mean value theorem, we have

$$|f(\mathbf{Y}+\mathbf{H})-f(\mathbf{Y})|\leq \max_{0\leq t\leq 1}|\langle \nabla f(\mathbf{Y}+t\mathbf{H},\mathbf{D}),\mathbf{H}\rangle|,$$

where $\langle\ ,\ \rangle$ is the standard scalar product on $(\mathcal{H}_n^{(\beta)})^p$. We have,

$$|f(\mathbf{Y}+\mathbf{H})-f(\mathbf{Y})|\leq \max_{0\leq t\leq 1}||\nabla f(\mathbf{Y}+t\mathbf{H},\mathbf{D})||_{\alpha'}||\mathbf{H}||_{\alpha}. \tag{2.37}$$

Note that if $m\in\mathbb{C}\langle \mathbf{X},\mathbf{D}\rangle$ is a monomial of degree k in $\mathbf{X}$ and l in $\mathbf{D}$, we can write $m=X_{i_1}^{n_1}D_{j_i}^{m_1}...X_{i_s}^{n_s}D_{j_s}^{m_s}$, with $\sum_j n_j=k$ and $\sum_i j_i=l$. Let $\mathbf{Z}=\mathbf{Y}+t\mathbf{H}$. We get by the non-commutative Hölder inequality (see [3, A 13]),

$$\mathrm{tr}|m(\mathbf{Z},\mathbf{D})|^{\alpha'}\leq \rho^{\alpha' l}\prod_{j=1}^{s}\left(\mathrm{tr}|Z_{i_j}|^{\alpha' k}\right)^{\frac{n_i}{k}}.$$

The arithmetico-geometric mean inequality yields,

$$\mathrm{tr}|m(\mathbf{Z},\mathbf{D})|^{\alpha'}\leq \rho^{\alpha' l}\sum_{j=1}^{s}\frac{n_i}{k}\mathrm{tr}|Z_{i_j}|^{\alpha' k}\leq \rho^{\alpha' l}\sum_{i=1}^{p}\mathrm{tr}|Z_i|^{\alpha' k}. \tag{2.38}$$

As $\nabla_{X_i}f$ is the sum of at most d monomials of degree $d-1$ in $\mathbf{X}$, and q in $\mathbf{D}$, we get by triangular inequality and the above observation,

$$||\nabla_{X_i}f(\mathbf{Z},\mathbf{D})||_{\alpha'}\leq d\rho^q||\mathbf{Z}||^{d-1}_{\alpha'(d-1)}.$$

Thus,

$$||\nabla f(\mathbf{Z},\mathbf{D})||_{\alpha'}\leq pd\rho^q||\mathbf{Z}||^{d-1}_{\alpha'(d-1)}.$$

As $\mathbf{X}\mapsto ||\mathbf{X}||^{d-1}_{\alpha'(d-1)}$ is convex and as $p\leq d$, we get as $\mathbf{Z}=\mathbf{Y}+t\mathbf{H}$,

$$||\nabla f(\mathbf{Z},\mathbf{D})||_{\alpha'}\leq d^2\rho^q(1+t)^{d-2}\big(||\mathbf{Y}||^{d-1}_{\alpha'(d-1)}+t||\mathbf{H}||^{d-1}_{v(d-1)}\big).$$

As $\alpha'(d-1)\geq\alpha$, we have

$$||\nabla f(\mathbf{Z},\mathbf{D})||_{\alpha'}\leq d^2\rho^q 2^{d-2}\big(||\mathbf{Y}||^{d-1}_{\alpha'(d-1)}+||\mathbf{H}||^{d-1}_{\alpha}\big).$$

This inequality together with (2.37) yields the claim (2.6.4). □

We come back now to the proof of Proposition 2.6.1.

Denote by m_f the median of $f(\mathbf{X})$, and let $A=\{f\leq m_f\}$. Define

$$B=\{\mathbf{Y}\in(\mathcal{H}_n^{(\beta)})^p,\ ||\mathbf{Y}||^{d-1}_{\alpha'(d-1)}\leq\kappa L_\alpha, ||\mathbf{Y}||^{d-1}_{2(d-1)}\leq\kappa L_2\}.$$

From the assumption (2.34) and Markov's inequality, we deduce that

$$\mathbb{P}(\mathbf{X}\in A\cap B)\geq\frac{1}{c_0}.$$

We have by Lemma 2.6.4 for any $\mathbf{Y}\in A'$, and $\mathbf{H}\in\kappa\sqrt{r}B_{\ell^2}$,

$$|f(\mathbf{Y}+\mathbf{H})-f(\mathbf{Y})|\leq C\rho^q\kappa^d\big(L_2\sqrt{r}+r^{d/2}\big). \tag{2.39}$$

Let $\mathbf{K} \in \kappa r^{1/\alpha} B_{\ell^\alpha}$. By Lemma 2.6.4 we get,

$$|f(\mathbf{Y}+\mathbf{H}+\mathbf{K}) - f(\mathbf{Y}+\mathbf{H})| \leq C\rho^q (||\mathbf{Y}+\mathbf{H}||^{d-1}_{\alpha'(d-1)} + ||\mathbf{K}||^{d-1}_{\alpha}\Big) ||\mathbf{K}||_\alpha,$$

with some $C > 0$ which will vary along the proof. Using the convexity of $\mathbf{Z} \mapsto ||\mathbf{Z}||^{d-1}_{k(d-1)}$, and the fact that $\alpha'(d-1) \geq \alpha$, we get

$$\begin{aligned} ||\mathbf{Y}+\mathbf{H}||^{d-1}_{\alpha'(d-1)} &\leq 2^{d-1}\Big(||\mathbf{Y}||^{d-1}_{\alpha'(d-1)} + ||\mathbf{H}||^{d-1}_{\alpha'(d-1)}\Big) \\ &\leq 2^{d-1}\Big(||\mathbf{Y}||^{d-1}_{\alpha'(d-1)} + ||\mathbf{H}||^{d-1}_{2}\Big) \end{aligned}$$

Therefore,

$$|f(\mathbf{Y}+\mathbf{H}+\mathbf{K}) - f(\mathbf{Y}+\mathbf{H})| \leq C\rho^q \Big(||\mathbf{Y}||^{d-1}_{\alpha'(d-1)} + ||\mathbf{H}||^{d-1}_{2} + ||\mathbf{K}||^{d-1}_{\alpha}\Big) ||\mathbf{K}||_\alpha, \quad (2.40)$$

As $\alpha \leq 2$, $B_{\ell^\alpha} \subset B_\alpha$ (see [104, Theorem 3.32]), we get

$$|f(\mathbf{Y}+\mathbf{H}+\mathbf{K}) - f(\mathbf{Y}+\mathbf{H})| \leq C\rho^q\kappa^d (L_\alpha r^{\frac{1}{\alpha}} + r^{\frac{d-1}{2}+\frac{1}{\alpha}} + r^{\frac{d}{\alpha}}). \quad (2.41)$$

We deduce from (2.39) and (2.41) that,

$$|f(\mathbf{Y}+\mathbf{H}+\mathbf{K}) - f(\mathbf{Y})| \leq 6C\rho^q\kappa^d \max\Big(L_2\sqrt{r}, L_\alpha r^{\frac{1}{\alpha}}, r^{\frac{d}{\alpha}}, r^{\frac{d-1}{2}+\frac{1}{\alpha}}, r^{\frac{d}{2}}\Big).$$

As $\alpha \leq 2$, we get

$$\max\Big(r^{\frac{d}{\alpha}}, r^{\frac{d-1}{2}+\frac{1}{\alpha}}, r^{\frac{d}{2}}\Big) = \max\Big(r^{\frac{d}{2}}, r^{\frac{d}{\alpha}}\Big).$$

We deduce that,

$$A' + \kappa\sqrt{r}B_{\ell^2} + \kappa r^{1/\alpha}B_{\ell^\alpha} \subset f^{-1}\big(-\infty, m_f + 6C\rho^q\kappa^d \max(L_2\sqrt{r}, L_\alpha r^{\frac{1}{\alpha}}, r^{\frac{d}{\alpha}}, r^{\frac{d}{2}})\big].$$

Thus, as $\mathbf{X}$ satisfies $\mathcal{C}_\alpha$,

$$\mathbb{P}(f > m_f + 6C\rho^q\kappa^d \max(L_2\sqrt{r}, L_\alpha r^{\frac{1}{\alpha}}, r^{\frac{d}{2}}, r^{\frac{d}{\alpha}}) \leq c_0 e^{-r},$$

which we can re-write as for any $t > 0$,

$$\mathbb{P}(f > m_f + t\kappa^d\rho^q) \leq c_0 \exp\Big(-c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}, t^{\frac{2}{d}}\Big)\Big),$$

with $c > 0$ a constant. But, as $\alpha \leq 2$, and $L_2 \geq 1$,

$$\min(t^{\frac{\alpha}{d}}, t^{\frac{2}{d}}) \geq \min\Big(t^{\frac{\alpha}{d}}, \frac{t^2}{L_2^2}\Big).$$

Therefore,

$$\mathbb{P}(f > m_f + t\kappa^d\rho^q) \leq c_0 \exp\Big(-c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}\Big)\Big).$$

The same argument carried out with $-f$ yields,

$$\mathbb{P}(|f - m_f| > t\kappa^d\rho^q) \leq 2c_0 \exp\Big(-c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}\Big)\Big), \quad (2.42)$$

Integrating this inequality gives,

$$|\mathbb{E}f(\mathbf{X}) - m_f| = O(\kappa^d \rho^q (L_2 + L_\alpha)).$$

We deduce for any $t > 0$,

$$\mathbb{P}(|f - \mathbb{E}f(\mathbf{X})| > t\kappa^d \rho^q + \delta_n) \leq 2c_0 \exp\Big(- c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}\Big)\Big),$$

with $\delta_n = O(\kappa\rho^q(L_2 + L_\alpha))$. Now, if $r \leq d$, observe that as L_2 and $L_\alpha \geq 1$, we have

$$||\mathbf{X}||_{2(r-1)}^{r-1} \leq L_2, \; ||\mathbf{X}||_{\alpha'(r-1)}^{r-1} \leq L_\alpha.$$

Thus, the same argument for $f = \mathrm{tr}Q$ with Q of degree $r \leq d$ in $\mathbf{X}$, entails

$$\mathbb{P}(|f - \mathbb{E}f(\mathbf{X})| > t\kappa\rho^q) \leq 2c_0 \exp\Big(- c\min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{r}}\Big)\Big).$$

As $r \leq d$, and $L_2, L_\alpha \geq 1$, we have

$$\min\big(t^{\frac{\alpha}{r}}, t^{\frac{2}{r}}\big) \geq \min\Big(\frac{t^2}{L_2^2}, \frac{t^\alpha}{L_\alpha^\alpha}, t^{\frac{\alpha}{d}}\Big).$$

Thus if $P \in \mathbb{C}\langle \mathbf{X}, \mathbf{D}\rangle$ is a polynomial of degree d in $\mathbf{X}$, we get the claim of the Proposition 2.6.1 by a union bound. $\square$

2.7 Conclusion and perspectives

We reviewed in this chapter concentration inequalities which allow to capture the large deviations behavior of the empirical spectral measure, the largest eigenvalue and traces of polynomials of Wigner matrices satisfying the concentration property $\mathcal{C}_\alpha$. We observed a grading of speeds for each of these functionals which illustrates the large deviations principles for Wigner matrices without Gaussian tails we presented in the introduction (see Theorems 1.8.4, 1.9.2, and 1.9.5). We saw that for what concern the traces of polynomials, the concentration inequality of Proposition 2.6.3 manages to capture the Gaussian fluctuations.

A further legitimate question is to ask if one can get concentration inequalities which capture the fluctuations of the spectral measure or the largest eigenvalue. With Proposition 2.3.1 (with the notable exception when $\alpha < 1$) we manage to capture the Gaussian fluctuations of

$$n\big(\mu_{X/\sqrt{n}}(f) - \mathbb{E}\mu_{X/\sqrt{n}}(f)\big),$$

when f is Lipschitz (by (2.13)). What could be said if f is not so regular? This limitation to Lipschitz functions is a consequence of the strategy we adopted which consists in considering $\mu_{X/\sqrt{n}}(f)$ as a function of the entries, and therefore imposes a kind of regularity to the test function f.

This obstacle, that is the fact that we are unable to derive concentration properties of the spectrum without using its stability with respect to the entries of the matrix, is also at the heart of the problem of getting concentration inequalities in the setting of non-normal matrices. Indeed, the spectrum of non-normal matrices is known to be very instable, as one sees with the example of the matrix N_t of size n,

$$N_t = \begin{pmatrix} 0 & \cdots & 0 & t \\ 1 & \ddots & & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots 0 & 1 & 0 \end{pmatrix},$$

for which, N_0 has spectrum $\{0\}$ and N_t, $\{t^{\frac{1}{n}} e^{\frac{2ik\pi}{n}} : k = 0, ..., n-1\}$ for $t > 0$. But if one considers for example a matrix with i.i.d coefficients with absolutely continuous laws with respect to the Lebesgue measure, then the matrix is almost surely diagonalizable, so that the situation described in the example above, does not actually happen. Some spectral variation inequalities are known for non-normal diagonalizable matrices, as

$$\mathcal{W}_2(\mu_A, \mu_B) \leq (c(P)c(Q))^{1/2} ||A - B||_2,$$

where $c(P) = ||P||\,||P^{-1}||$, and similarly for $c(Q)$, are the condition numbers of the transition matrices to the of A and B respectively (see [22, Theorem VIII.3.10]). But they are not tight, and therefore this gives hope for some improvement in that direction.

Alternatively, attacking this problem from an opposite direction, we mention the work of Chafaï, Hardy and Maïda [37] who established concentration inequalities for the Ginibre ensemble using the determinantal form of the joint law of the eigenvalues.

As for the largest eigenvalue, it remains quite a challenge to get a concentration inequality which reflects the convergence in law of the largest eigenvalue of Wigner matrices to the Tracy-Widom law. The only sharp inequalities we are aware of, are due to Ledoux in [71], for the classical Gaussian ensembles, as the GUE, GOE or the Laguerre ensemble, using semi-groups techniques, and to Ledoux and Rider for case of the β-ensembles [72].

3. Large deviations for Wigner matrices without Gaussian tails

This chapter is based on the article *Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails*, Electron. J. Probab. **21** (2016), no 32, 1-49.

3.1 Introduction

We saw in the preceding chapter, that the speed of deviations of the empirical spectral measure, the largest eigenvalue or the trace of polynomials of Wigner matrices is very sensitive to the concentration property of the entries, in the regime where this property is weaker than normal concentration. This situation, at least for the empirical spectral measure, is in strong contrast with the classical setting, where the speed of the deviations of the empirical measure of an i.i.d sample of size n is universal by Hoeffding's inequality (see [76, Chapter 2 §2.6]), and for which we have a full LDP with speed n by Sanov's theorem [43, Theorem 6.2.10]. But this change of speed is also the sign of a heavy-tail phenomenon appearing in the deviations of the spectrum, meaning that only large entries or large eigenvalues are participating in the deviations, when the law of the matrix has concentration $\mathcal{C}_\alpha$ for $\alpha < 2$.

Taking advantage of this phenomenon, Bordenave and Caputo derived a full LDP for the empirical spectral measure of the so-called model of Wigner matrices without Gaussian tails, which they introduced in [29]. Their approach turned out to be very efficient in understanding the large deviations of spectral functionals of this model. In the setting of Wishart matrices, a LDP is known for the empirical spectral measure due to Groux [55], whereas for our part, we obtained LDPs for the largest eigenvalue and traces of powers of Wigner matrices without Gaussian tails. We will present in this chapter the LDP for the largest eigenvalue.

3.2 Main results

We recall the model of Wigner matrices without Gaussian tails, with which we will be working in this chapter.

3.2.1 Definition. We say that a Wigner matrix X is *without Gaussian tails* if there exist $\alpha \in (0,2)$ and $a, b \in (0,+\infty)$ such that,

$$\lim_{t\to+\infty} -t^{-\alpha} \log \mathbb{P}\left(|X_{1,1}| > t\right) = b, \tag{3.1}$$

$$\lim_{t\to+\infty} -t^{-\alpha} \log \mathbb{P}\left(|X_{1,2}| > t\right) = a,$$

and there are two probability measures on $\mathbb{S}^1$, υ_1 and υ_2, and $t_0 > 0$, such that for all $t \geq t_0$ and any measurable subset U of $\mathbb{S}^1$,

$$\mathbb{P}\left(X_{1,1}/|X_{1,1}| \in U, |X_{1,1}| \geq t\right) = \upsilon_1(U)\mathbb{P}\left(|X_{1,1}| \geq t\right),$$

$$\mathbb{P}\left(X_{1,2}/|X_{1,2}| \in U, |X_{1,2}| \geq t\right) = \upsilon_2(U)\mathbb{P}\left(|X_{1,2}| \geq t\right). \tag{3.2}$$

Concerning the largest eigenvalue of Wigner matrices without Gaussian tails, we obtained the following result.

3.2.2 Theorem. *Let X be a Wigner matrix without Gaussian tails. We assume that $\Re(X_{1,2})$ and $\Im(X_{1,2})$ are independent.The sequence $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ follows a LDP with speed $n^{\alpha/2}$, and good rate function defined for all $x \in \mathbb{R}$, by*

$$J_\alpha(x) = \begin{cases} c g_{\mu_{sc}}(x)^{-\alpha} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2, \end{cases}$$

where c is a constant depending only on α, a and b, and where $g_{\mu_{sc}}$ denotes the Stieltjes transform of the semicircular law.

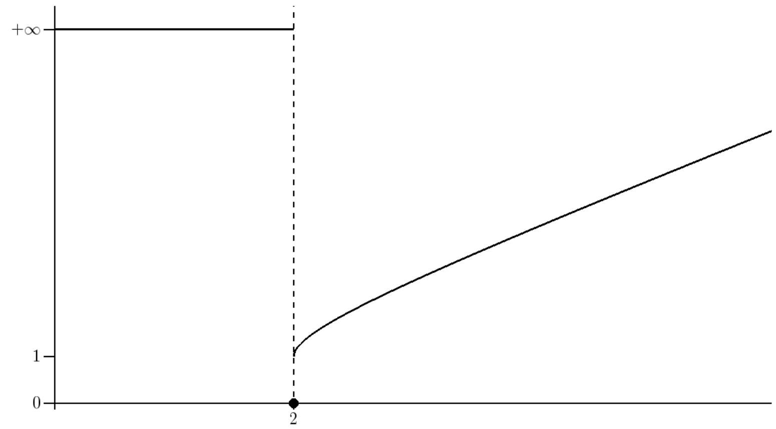
3.2.3 Remark. The assumption on the independence of the real and imaginary parts of the off-diagonal entries is purely technical. We only make this assumption in order to use the estimates in [81] on the entries of the resolvent, which we will need in our proof of the LDP. In particuliar, it is not needed for the LDP of the empirical spectral measure in [29].

We will prove that the constant c in Theorem 3.2.2, can be computed explicitly in certain cases, in particular when the entries are real random variables. We refer the reader to the Section 3.11 for further details.

Observe that the rate function is infinite on $(-\infty, 2)$. Indeed, in order to make a deviation of the top eigenvalue at the left of 2, we need to force the support of the empirical spectral measure to be in $(-\infty, 2-\varepsilon)$, for some $\varepsilon > 0$. But this event has an infinite cost at the exponential scale $n^{\alpha/2}$ since the spectral measure follows a LDP with speed $n^{1+\alpha/2}$.

As illustrated in figure 3.1, drawn in the case $\alpha = 1$, this rate function is discontinuous at 2. As we will show, the deviations of the top eigenvalue are given by finite rank perturbations of a Wigner matrix. As we mentioned in the Introduction §1.5, the behavior of the extreme eigenvalues of these deformed models are known to present a threshold phenomenon with respect to the strength of the perturbation, which the rate function reflects through its discontinuity at 2. This picture may also mean that there is a more subtle behavior of the largest eigenvalue in the right neighborhood of 2, which is to be understood.

Figure 3.1: Graph of the rate function J



3.3 Heuristics

We will show that one can obtain the lower bound of the LDP by a finite rank perturbation. For simplicity, let us assume that the $X_{i,j}$'s are exponential variables with parameter 1. Thus, the matrix X satisfies the assumptions (3.2.1) with $\alpha = 1$, and $a = b = 1$. In this case, Proposition 3.11.1 shows that the constant c in Theorem 3.2.2 is 1.

Let $x > 2$ and $\theta = 1/g_{\mu_{sc}}(x)$. As $g_{\mu_{sc}}$ maps $[2, +\infty)$ into $(0, 1]$, we have $\theta > 1$. By independence of the entries, we have

$$\mathbb{P}\big(\lambda_{X/\sqrt{n}} \simeq x\big) \gtrsim \mathbb{P}\big(\lambda_{X^{(1)}/\sqrt{n}+\theta e_1 e_1^*} \simeq x\big)\mathbb{P}\big(X_{1,1}/\sqrt{n} \simeq \theta\big), \tag{3.3}$$

with $X^{(1)} = X - X_{1,1}e_1e_1^*$, and e_1 the first coordinate vector of $\mathbb{C}^n$. Since $\theta > 1$, we have according to [82],

$$\lambda_{X/\sqrt{n}+\theta e_1 e_1^*} \underset{n\to+\infty}{\longrightarrow} g_{\mu_{sc}}^{-1}(1/\theta),$$

in probability. Using Weyl's inequality (1.6) we get, recalling that $x = g_{\mu_{sc}}^{-1}(1/\theta)$,

$$\mathbb{P}\big(\lambda_{X^{(1)}/\sqrt{n}+\theta e_1 e_1^*} \simeq x\big) \underset{n\to+\infty}{\longrightarrow} 1. \tag{3.4}$$

But $X_{1,1}$ has exponential law with parameter 1, thus

$$\mathbb{P}\big(X_{1,1}/\sqrt{n} \simeq \theta\big) \simeq e^{-\theta\sqrt{n}}. \tag{3.5}$$

Putting together (3.3), (3.4) and (3.5), we get,

$$\mathbb{P}\big(\lambda_{X/\sqrt{n}} \simeq x\big) \gtrsim e^{-g_{\mu_{sc}}(x)^{-1}\sqrt{n}}.$$

which is the lower bound expected by Theorem 3.2.2 and Proposition 3.11.1, for $\alpha = 1$ and $a = b = 1$. Note that we could also have used a deformation of the type

$$\begin{pmatrix} 0 & \theta \\ \theta & 0 \end{pmatrix},$$

to get the lower bound of the LDP.

3.4 Outline of proof

The strategy of the proof will closely follow the one of the LDP for the empirical spectral measure derived in [29].

Following [29], we start by cutting the entries of X_N according to their size. We decompose X in the following way. Fix some $d > 0$ such that $d\alpha > 1$, and let $\varepsilon > 0$. We write,

$$X/\sqrt{n} = A + B^{\varepsilon} + C^{\varepsilon} + D^{\varepsilon}, \tag{3.6}$$

with, for all $i, j \in \{1, ..., n\}$,

$$A_{i,j} = \mathbb{1}_{|X_{i,j}|_\infty \leq (\log n)^d} \frac{X_{i,j}}{\sqrt{n}}, \qquad B^{\varepsilon}_{i,j} = \mathbb{1}_{(\log n)^d < |X_{i,j}|_\infty < \varepsilon n^{1/2}} \frac{X_{i,j}}{\sqrt{n}},$$

$$C^{\varepsilon}_{i,j} = \mathbb{1}_{\varepsilon n^{1/2} \leq |X_{i,j}|_\infty \leq \varepsilon^{-1} n^{1/2}} \frac{X_{i,j}}{\sqrt{n}}, \qquad D^{\varepsilon}_{i,j} = \mathbb{1}_{\varepsilon^{-1} n^{1/2} < |X_{i,j}|_\infty} \frac{X_{i,j}}{\sqrt{n}},$$

where $|z|_\infty = \max(|\Re(z)|, |\Im(z)|)$ for all complex numbers z.

Our first step will be to prove some concentration inequalities in Section 3.5, which we will use throughout this chapter, and in particular to prove the exponential tightness of $(\lambda_{X\sqrt{n}})_{n\in\mathbb{N}}$ in Section 3.6.

Then, in Section 3.7, we will focus on trying to identify which parts in the decomposition of $X/\sqrt{n}$ significantly contribute to create deviations of the largest eigenvalue. We start by showing in Section 3.7.1, that we can neglect the contributions of B^{ε} and D^{ε}, corresponding to the intermediate and large entries respectively, in the deviations of $\lambda_{X/\sqrt{n}}$. Then in Section 3.7.2, we prove that we can replace A by a Hermitian matrix H_n, with entries bounded by $(\log n)^d/\sqrt{n}$, and independent from C^{ε}.

From the LDP of the empirical spectral measure of $X/\sqrt{n}$ of speed $n^{1+\alpha/2}$ proved in [29], we deduce in Proposition 3.8.1 that the deviations at the left of 2 have an infinite cost at the scale $n^{\alpha/2}$. Therefore, we only need to focus on the deviations of the largest eigenvalue of $H_n + C^{\varepsilon}$ at the right of 2. As in many papers on finite rank deformations of Wigner matrices (see [20] for exemple), we see the largest eigenvalue of $H_n + C^{\varepsilon}$, provided it is not in the spectrum of H_n, as the largest zero of the function f_n defined outside the spectrum of H_n by,

$$f_n(x) = \det(M_n(x)), \text{ with } M_n(x) = I_k - (\theta_i \langle u_i, (x - H_n)^{-1} u_j \rangle)_{1 \leq i,j \leq k},$$

where k is the rank of C^ε, $\theta_1, ..., \theta_k$ are the non-zero eigenvalues of C^ε in non-decreasing order, and $u_1, ..., u_k$ are orthonormal eigenvectors of C^ε associated to $\theta_1, ..., \theta_k$.

As we will see, this method is made efficient in the study of the deviations of $\lambda_{H_n+C^\varepsilon}$ at the right of 2 by two main facts. Firstly, as we show in Proposition 3.8.3, the spectrum of H_n can be considered at the exponential scale $n^{\alpha/2}$ nearly as contained in $(-\infty, 2]$. Secondly, as shown in Lemma 3.6.7, C^ε is a sparse matrix so that its rank can be considered at the exponential scale $n^{\alpha/2}$ as bounded.

In Section 3.8, we focus on showing that the function f_n is exponentially equivalent to a certain limit function f, defined for any $x > 2$ by,

$$f(x) = \prod_{i=1}^{k} (1 - \theta_i g_{\mu_{sc}}(x)) .$$

To this end, we show in Proposition 3.8.6, using concentration inequalities, that at the exponential scale $n^{\alpha/2}$, and uniformly in x in a compact subset of $(2, +\infty)$,

$$M_n(x) \simeq I_k - (\theta_i \langle u_i, \mathbb{E}\,(x - H_n)^{-1}\, u_j\rangle)_{1\le i,j\le k}. \tag{3.7}$$

(We use in the proof some caracteristic function restraining the spectrum of H_n, to make sense of the expectation on the right-hand side). Next, in Theorem 3.8.7, we prove an isotropic property of the semi-circular law using the estimates in [81] of the entries of the resolvent of Wigner matrices. This allows us to deduce in Proposition 3.8.8 that

$$M_n(x) \simeq I_k - \begin{pmatrix} \theta_1 g_{\mu_{sc}}(x) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \theta_k g_{\mu_{sc}}(x) \end{pmatrix}.$$

Using the fact that the spectral radius of C^ε can be considered as bounded as shown in Lemma 3.6.5, and using the uniform continuity of the determinant on compact sets of $\mathcal{H}_k^{(\beta)}$, we get, as stated in Theorem 3.8.4, uniformly in x in any compact subset contained in $(2, +\infty)$,

$$f_n(x) \simeq f(x), \text{ with } f(x) = \prod_{i=1}^{k} (1 - \theta_i g_{\mu_{sc}}(x)) .$$

In Section 3.9, we show that provided $\lambda_{H_n+C^\varepsilon}$ is greater that 2, and that λ_{C^ε} is greater than 1, the largest zero of f_n, namely $\lambda_{H_n+C^\varepsilon}$, is exponentially equivalent to the largest zero $\rho_{n,\varepsilon}$ of f.

Despite the fact that f_n and f are holomorphic functions, we cannot use Rouché's theorem to deduce that their zeros are close since we only know that they are close on compact subsets of $(2, +\infty)$. We use here a trick a bit similar to the one used in [20, p. 513], which will allow us to make do with this uniform closeness between f_n and f on compact subsets of $(2, +\infty)$. We perturb the spectrum of

C^ε so that its largest eigenvalue is simple and bounded away from its second largest eigenvalue by some $\gamma > 0$. Classical intermediate values theorem then shows that any continuous function close to f on all compact subsets contained in $(2, +\infty)$, admits a zero in $(2, +\infty)$, and that its largest zero is close to the largest zero of f. Since f remains in a compact set of continuous functions, we can prove a uniform continuity property for the "largest zero function" in Lemma 3.9.3. In Proposition 3.9.2, we deduce that the largest zero of f_n and of f are exponentially equivalent at the scale $n^{\alpha/2}$. This allows us to conclude in Theorem 3.9.1 that $(\rho_{n,\varepsilon})_{n\in\mathbb{N},\varepsilon>0}$, are an exponentially good approximations of $\lambda_{X/\sqrt{n}}$ (in the sense of [43, Definition 4.2.10]).

Then, in Section 3.10, we prove that $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ satisfies a LDP for each $\varepsilon > 0$, and we deduce a LDP for $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$. The key of the proof is Proposition 3.6.7, which allows us to assume that the matrix C^ε has only a finite number of non-zero entries at the exponential scale $n^{\alpha/2}$. With this observation, the problem can be reduced to a finite-dimensional one. We define $\widetilde{\mathcal{E}}_r$ to be the set of equivalence classes of infinite Hermitian matrices with at most r non-zero entries, under the action of permutation matrices. In Proposition 3.10.1, we establish a LDP for C^ε, when seen as an element of $\widetilde{\mathcal{E}}_r$, with respect to the topology given by the distance,

$$\forall \tilde{A}, \tilde{B} \in \widetilde{\mathcal{E}}_r,\ \tilde{d}\left(\tilde{A}, \tilde{B}\right) = \min_{\sigma,\sigma'\in\mathfrak{S}} \max_{i,j} \left|B_{\sigma(i),\sigma(j)} - A_{\sigma'(i),\sigma'(j)}\right|,$$

where A and B representatives of $\tilde{A}$ and $\tilde{B}$ respectively, and where $\mathfrak{S} = \cup_{m\in\mathbb{N}}\mathfrak{S}_m$ is the union of the symmetric groups. The map which associates to any matrix of $\widetilde{\mathcal{E}}_r$, its largest eigenvalue is continuous with respect to $\tilde{d}$, and allows us to apply a contraction principle to get the LDP for $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$, which is stated in Proposition 3.10.3. We finally deduce a LDP for $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ in Theorem 3.10.4, with rate function

$$J(x) = \begin{cases} c g_{\mu_{sc}}(x)^{-\alpha} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2, \end{cases}$$

where

$$c = \inf\Big\{ b\sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + a\sum_{i\neq j} |A_{i,j}|^\alpha : \lambda_A = 1, A \in \mathcal{D} \Big\}, \tag{3.8}$$

and

$$\mathcal{D} = \Big\{ A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)} : \forall i \leq j,\ A_{i,j} = 0 \text{ or } \frac{A_{i,j}}{|A_{i,j}|} \in \operatorname{supp}(\nu_{i,j}) \Big\},$$

where $\nu_{i,j} = \nu_1$ if $i = j$, and ν_2 if $i < j$, and where $\operatorname{supp}(\nu_{i,j})$ denotes the support of the measure $\nu_{i,j}$.

In Section 3.11, we show that we can compute explicitly in certain cases the constant c appearing in the rate function J. In particular, in the case where the entries of $X/\sqrt{n}$ are real, or when $\alpha \in (0, 1]$, Proposition 3.11.1 computes completely the constant c.

The optimization problem (3.8) exhibits two different behaviors, when $\alpha \in (0, 1]$ and when $\alpha \in (1, 2)$. When $\alpha \in (0, 1]$, the infimum is achieved for matrices of sizes

1 or 2, and can computed for any choice of ν_1 and ν_2. When $\alpha \in (1,2)$, the picture is more complicated, and one cannot say much without some assumptions on the supports of ν_1 and ν_2. In particular, one can observe that when $b > \frac{a}{2}$ and $1 \in \mathrm{supp}(\nu_1) \cap \mathrm{supp}(\nu_2)$, the infimum can be achieved for a matrix of size arbitrary large, when α gets arbitrary close to 2.

Moreover, the knowledge of the minimizers of (3.8) is useful to derive the lower bound of the LDP. Indeed, it indicates which finite rank deformation one has to choose to get the lower bound on the deviations of $\lambda_{X/\sqrt{n}}$, as explained in Section 3.3.

3.5 Concentration inequalities

Throughout the rest of the chapter, we fix a constant $\kappa > 0$, such that for all t large enough,

$$\mathbb{P}\left(|X_{1,1}| > t\right) \vee \mathbb{P}\left(|X_{1,2}| > t\right) \leq e^{-\kappa t^{\alpha}}. \tag{3.9}$$

With a slight adaptation of the concentration inequality from [76, Chapter 8 Example 8.7], for the largest eigenvalue of a random symmetric matrix with bounded entries, we get the following proposition.

3.5.1 Proposition. *Let H be a random Hermitian matrix with entries bounded by a constant $K > 0$, such that $(H_{i,j})_{i \leq j}$ are independent variables and let C be a deterministic Hermitian matrix. For all $t > 0$,*

$$\mathbb{P}\left(|\lambda_{H+C} - \mathbb{E}\lambda_{H+C}| > t\right) \leq 2\exp\Big(-\frac{t^2}{32K^2}\Big).$$

We state now a local concentration inequality (see [56, Chapter 5 §5.4]) we will use later in order to prove an isotropic-like property of the semi-circle law.

3.5.2 Proposition. *Let u be a unit vector of $\mathbb{C}^n$, and $\mu \in \mathbb{R}$. Let H be a random Hermitian matrix of size n, such that the entries $(H_{i,j})_{1 \leq i \leq j \leq n}$ are independent and bounded by $K > 0$. We denote by $\mathcal{C}$, the set of Hermitian matrices X of size n, with top eigenvalue λ_X strictly less that μ. Let also $x \in (\mu, +\infty)$.*

(i). The function $f_u : \mathcal{C} \to \mathbb{R}$ defined by

$$f_u(X) = \langle u, (x - X)^{-1} u \rangle,$$

is convex and $1/(x-\mu)^2$-Lipschitz with respect to the Hilbert-Schmidt norm $|| \, ||_2$.

(ii). f_u admits a convex extension to $\mathcal{H}_n^{(\beta)}$, denoted $\tilde{f}_u$ which is $1/(x-\mu)^2$-Lipschitz with respect to the Hilbert-Schmidt norm.
Moreover, for all $x > \mu$, and all $t > 0$,

$$\mathbb{P}\big(\big|\tilde{f}_u(H) - \mathbb{E}(\tilde{f}_u(H))\big| > t\big) \leq 2\exp\Big(-\frac{(x-\mu)^4 t^2}{32K^2}\Big).$$

Proof. (i). Let $x > \mu$. From [22, Exercice V.1.15], we know that $t \mapsto 1/t$ is operator convex on $(0, +\infty)$, meaning that on the set of positive matrices $A \mapsto A^{-1}$ is convex for the matrix order. Consequently, $t \mapsto (x-t)^{-1}$ is operator convex on $(-\infty, x)$, and in particular on $(-\infty, \mu)$. It means that the mapping f_u, defined on $\mathcal{C}$ by,

$$f_u(X) = \langle u, (x-X)^{-1} u \rangle,$$

is convex. Since $x > \mu$, we have for all X, Y in $\mathcal{C}$,

$$\begin{aligned} f_u(X) - f_u(Y) &= \langle u, \big((x-X)^{-1} - (x-Y)^{-1}\big) u \rangle \\ &= \langle u, (x-X)^{-1} (X-Y) (x-Y)^{-1} u \rangle \\ &\leq \frac{1}{(x-\mu)^2} \left|\left| X - Y \right|\right|_2 . \end{aligned}$$

Thus, f_u is convex and $1/(x-\mu)^2$-Lipschitz.

(ii). Since f_u is differentiable, we can write for all $X \in \mathcal{C}$

$$f_u(X) = \sup_{Y \in \mathcal{C}} \left(f_u(Y) + \langle \nabla f_u(Y), (X-Y) \rangle \right),$$

where $\langle \ , \ \rangle$ denotes the canonical Hermitian product on the space of Hermitian matrices of size n, denoted $\mathcal{H}_n^{(\beta)}$. Let $\widetilde{f}_u$ be defined for all $X \in \mathcal{H}_n^{(\beta)}$ by

$$\widetilde{f}_u(X) = \sup_{Y \in \mathcal{C}} \left(f_u(Y) + \langle \nabla f_u(Y), (X-Y) \rangle \right).$$

For all $X \in \mathcal{H}_n^{(\beta)}$, $\tilde{f}_u(X) < +\infty$, since for all $Y \in \mathcal{C}$,

$$\left|\left| \nabla f_u(Y) \right|\right|_2 \leq \frac{1}{(x-\mu)^2}.$$

As a supremum of affine functions, $\widetilde{f}_u$ is convex and by the property above it is also $1/(x-\mu)^2$-Lipschitz.

We show now that $\widetilde{f}_u$ satisfies a bounded differences inequality in quadratic mean, in the sense of [76, Theorem 8.6], on the product space $\mathcal{H}_n^{(\beta)}$ of Hermitian matrices with entries bounded by K. Let H and H' be two Hermitian matrices with entries bounded by K. Let $\zeta(H)$ be a sub-differential of $\widetilde{f}_u$ at the point H. Then we have,

$$\begin{aligned} \tilde{f}_u(H) - \tilde{f}_u(H') &\leq \langle \zeta(H), (H - H') \rangle \\ &\leq \sum_{1 \leq i \leq j \leq n} \mathbb{1}_{H_{i,j} \neq H'_{i,j}} 4K |\zeta(H)_{i,j}|, \end{aligned}$$

where $\zeta(H)_{i,j}$ denote the (i,j) coordinate of $\zeta(H)$. Since $\tilde{f}_u$ is $1/(x-\mu)^2$-Lipschitz we have,

$$\left|\left| \zeta(H) \right|\right|_2 \leq \frac{1}{(x-\mu)^2}.$$

Using [76, Theorem 8.6], it follows that for all $t > 0$,

$$\mathbb{P}\big(|\widetilde{f}_u(H) - \mathbb{E}(\widetilde{f}_u(H))| > t\big) \leq 2 \exp\big(- \frac{(x-\mu)^4 t^2}{32K^2} \big).$$

$\square$

3.6 Exponential tightness

The goal of this section is to prove that $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ is exponentially tight at the exponential scale $n^{\alpha/2}$. More precisely, we will prove the following.

3.6.1 Proposition.

$$\lim_{t\to+\infty}\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}(\lambda_{X/\sqrt{n}}>t)=-\infty.$$

Proof. According to Weyl's inequality (1.6) we have,

$$\lambda_{X/\sqrt{n}}\le\lambda_A+\lambda_{B^\varepsilon}+\lambda_{C^\varepsilon}+\lambda_{D^\varepsilon},$$

where A, B^ε, C^ε, and D^ε are as in (3.6). Therefore

$$\begin{aligned}\mathbb{P}\left(\lambda_{X_N}>4t\right)&\le\mathbb{P}\left(\lambda_A>t\right)+\mathbb{P}\left(\lambda_{B^\varepsilon}>t\right)\\&\quad+\mathbb{P}\left(\lambda_{C^\varepsilon}>t\right)+\mathbb{P}\left(\lambda_{D^\varepsilon}>t\right).\end{aligned}\tag{3.10}$$

We are going to estimate at the exponential scale $N^{\alpha/2}$ the probability of each of the events $\{\lambda_A>t\}$, $\{\lambda_{B^\varepsilon}>t\}$, $\{\lambda_{C^\varepsilon}>t\}$, and $\{\lambda_{D^\varepsilon}>t\}$.

From the assumption (3.2.1) on the tail distributions of the entries, we get the following lemma, which we state without proof.

3.6.2 Lemma. *For $t>0$,*

$$\mathbb{E}\left(\mathbb{1}_{|X_{1,1}|>t}|X_{1,1}|^2\right)\vee\mathbb{E}\left(\mathbb{1}_{|X_{1,2}|>t}|X_{1,2}|^2\right)=O\left(e^{-\frac{\kappa}{2}t^\alpha}\right),$$

with $\kappa>0$ as in (3.9).

We focus first on the event $\{\lambda_A>t\}$. Applying the result of Proposition 3.5.1, we get the following corollary.

3.6.3 Corollary. *For all $t>0$,*

$$\lim_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\left(|\lambda_A-2|>t\right)=-\infty,\tag{3.11}$$

where A is as in (3.6).

Proof. If we apply Proposition 3.5.1 to A, with $K=\frac{(\log n)^d}{\sqrt{n}}$ we get for any $t>0$,

$$\mathbb{P}\left(|\lambda_A-\mathbb{E}\lambda_A|>t/2\right)\le 2\exp\Big(-\frac{t^2n}{128(\log n)^{2d}}\Big).$$

Since $\alpha<2$, we have

$$\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\left(\left|\lambda_A-\mathbb{E}\left(\lambda_A\right)\right|>t/2\right)=-\infty.\tag{3.12}$$

We know from [3, Exercice 2.1.27] that the largest eigenvalue of $X/\sqrt{n}$ converges in mean to 2. Besides by Weyl's inequality (1.6) we have,

$$\begin{aligned}\mathbb{E}|\lambda_A-\lambda_{X/\sqrt{n}}|^2&\le\mathbb{E}\mathrm{tr}\Big(A-\frac{X}{\sqrt{n}}\Big)^2\\&=\frac{1}{N}\sum_{1\le i,j\le n}\mathbb{E}\big(|X_{i,j}|^2\,\mathbb{1}_{|X_{i,j}|>(\log n)^d}\big).\end{aligned}\tag{3.13}$$

But from Lemma 3.6.2 we have,

$$\mathbb{E}\big(\mathbb{1}_{|X_{i,j}|>(\log n)^d}\,|X_{i,j}|^2\big) = O\big(e^{-\frac{\kappa}{2}(\log n)^{d\alpha}}\big),$$

with $\kappa > 0$ defined in (3.9). Putting the estimate above into (3.13), we get together with the fact that $d\alpha > 1$,

$$\mathbb{E}|\lambda_A - \lambda_{X/\sqrt{n}}|^2 \underset{n\to+\infty}{\longrightarrow} 0,$$

which implies

$$\mathbb{E}\lambda_A \underset{n\to+\infty}{\longrightarrow} 2. \tag{3.14}$$

Putting together (3.12) and (3.14), we get

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_A - 2| > t\right) = -\infty.$$

□

We can deduce from Proposition 3.6.3 that for t large enough, we have,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(\lambda_A > t\right) = -\infty. \tag{3.15}$$

For the second event $\{\lambda_{B^\varepsilon} > t\}$, we start by proving the following lemma.

3.6.4 Lemma. *For all $t > 0$,*

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\mathrm{tr}\,(B^\varepsilon)^2 > t\big) \leq -\frac{2^{\alpha/2}}{8} t\kappa\alpha\varepsilon^{-2+\alpha},$$

with $\kappa > 0$ as in (3.9).

Proof. We repeat here almost verbatim the argument used in the proof of [29, Lemma 2.3]. We have,

$$\begin{aligned}\mathbb{P}(\mathrm{tr}\,(B^\varepsilon)^2 > t) &\leq \mathbb{P}\Big(2\sum_{i\leq j} \frac{|X_{i,j}|^2}{n} \mathbb{1}_{(\log n)^d < |X_{i,j}|_\infty < \varepsilon n^{1/2}} > t\Big)\\ &\leq \mathbb{P}\Big(\sum_{i\leq j} \frac{|X_{i,j}|^2}{n} \mathbb{1}_{(\log n)^d < |X_{i,j}| < \sqrt{2}\varepsilon n^{1/2}} > \frac{t}{2}\Big),\end{aligned}$$

where we used in the last inequality $|X_{i,j}|_\infty \leq |X_{i,j}| \leq \sqrt{2}|X_{i,j}|_\infty$.

Let now $\lambda > 0$. By Chernoff's inequality,

$$\mathbb{P}(\mathrm{tr}\,(B^\varepsilon)^2 > t) \leq e^{-\lambda\frac{t}{2}} \prod_{i\leq j} \mathbb{E}\Big(\exp\Big(\lambda \frac{|X_{i,j}|^2}{n} \mathbb{1}_{(\log n)^d < |X_{i,j}| < \sqrt{2}\varepsilon n^{1/2}}\Big)\Big). \tag{3.16}$$

We denote by $\Lambda_{i,j}$ be the Laplace transform of $\mathbb{1}_{(\log N)^d < |X_{i,j}| < \sqrt{2}\varepsilon N^{1/2}} |X_{i,j}|^2/\sqrt{n}$, and by μ the distribution of $|X_{i,j}|$. Then, we have

$$\Lambda_{i,j}(\lambda) \leq 1 + \int_{(\log n)^d}^{\sqrt{2}\varepsilon n^{1/2}} e^{\frac{\lambda x^2}{n}} \mathrm{d}\mu(x).$$

Recall that for μ a probability measure on $\mathbb{R}$, and $g \in C^1$, we have the following integration by parts formula:

$$\int_a^b g(x)d\mu(x) = g(a)\mu\left[a,+\infty\right) - g(b)\mu\left(b,+\infty\right) + \int_a^b g'(x)\mu\left[x,+\infty\right)dx. \tag{3.17}$$

Thus,

$$\Lambda_{i,j}(\lambda) \leq 1 + \mu[(\log n)^d, +\infty)e^{\frac{\lambda(\log n)^{2d}}{n}} + \int_{(\log n)^d}^{\sqrt{2}\varepsilon n^{1/2}} \frac{2\lambda x}{n} e^{\frac{\lambda x^2}{n}} \mu[x,+\infty)dx.$$

We define $f(x) = \frac{\lambda x^2}{n} - \kappa x^\alpha$, with κ as in (3.9). For n large enough we get,

$$\begin{aligned}\Lambda_{i,j}(\lambda) &\leq 1 + e^{f\left((\log n)^d\right)} + \int_{(\log n)^d}^{\sqrt{2}\varepsilon n^{1/2}} \frac{2\lambda}{n} x e^{f(x)}dx \\ &\leq 1 + e^{f\left((\log n)^d\right)} + 4\lambda\varepsilon^2 \max_{[(\log n)^d, \sqrt{2}\varepsilon n^{1/2}]} e^f.\end{aligned} \tag{3.18}$$

Choose $\lambda = 2^{\alpha/2-2}\kappa\alpha\varepsilon^{-2+\alpha}n^{\alpha/2}$. Observe that f is decreasing until x_0 and increasing on $[x_0, +\infty)$, with x_0 given by

$$x_0 = \left(\frac{\kappa\alpha n}{2\lambda}\right)^{1/(2-\alpha)} = \left(2^{1-\alpha/2}n^{1-\alpha/2}\varepsilon^{2-\alpha}\right)^{1/(2-\alpha)} = \sqrt{2}\varepsilon n^{1/2}.$$

Thus, the maximum of e^f on $[(\log n)^d, \sqrt{2}\varepsilon n^{1/2}]$ is achieved at $(\log n)^d$. Since $\alpha/2 < 1$, we have for n large enough,

$$f((\log n)^d) = 2^{\alpha/2-2}\kappa\alpha\varepsilon^{-2+\alpha}n^{\alpha/2-1}(\log n)^{2d} - \kappa(\log n)^{d\alpha} \leq -\frac{\kappa}{2}(\log n)^{d\alpha}.$$

From (3.18) and the inequality above, we get

$$\Lambda_{i,j}(\lambda) \leq 1 + e^{-\frac{\kappa}{2}(\log n)^{d\alpha}}\left(1 + 2^{\alpha/2}\kappa\alpha\varepsilon^\alpha n^{\alpha/2}\right).$$

Since $d\alpha > 1$, we have for N large enough

$$\Lambda_{i,j}(\lambda) \leq 1 + e^{-\frac{\kappa}{4}(\log n)^{d\alpha}} \leq \exp\left(e^{-\frac{\kappa}{4}(\log n)^{d\alpha}}\right).$$

Finally, putting this last estimate into (3.16) we get

$$\mathbb{P}(\mathrm{tr}(B^\varepsilon)^2 > t) \leq \exp\Big(-\frac{2^{\alpha/2}}{8}t\kappa\alpha\varepsilon^{-2+\alpha}n^{\alpha/2}\Big)\exp\left(n^2e^{-\frac{\kappa}{4}(\log n)^{d\alpha}}\right), \tag{3.19}$$

which gives the claim. □

Coming back at the proof of Proposition 3.6.1, we observe that

$$\mathbb{P}(\lambda_{B^\varepsilon} > t) \leq \mathbb{P}(\mathrm{tr}(B^\varepsilon)^2 > t^2).$$

Hence,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(\lambda_{B^\varepsilon} > t\right) \leq -\frac{2^{\alpha/2}}{8}t^2\kappa\alpha\varepsilon^{-2+\alpha}. \tag{3.20}$$

We focus now on the third event $\{\lambda_{C^\varepsilon} > t\}$. The estimate is given by the following lemma.

3.6.5 Lemma. *For all* $t > 0$,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(||C^\varepsilon|| > t\right) \leq -\frac{\kappa}{4\sqrt{2}} t\varepsilon^{\alpha+1}, \tag{3.21}$$

with κ *as in* (3.9) *and where* $||\ ||$ *denotes the operator norm.*

Proof. As

$$||C^\varepsilon|| \leq ||C^\varepsilon||_{\ell^1\to\ell^1} = \max_{1\leq i\leq n} \sum_{j=1}^n |C^\varepsilon_{i,j}|,$$

where $||\ ||_{\ell^1\to\ell^1}$ denotes the operator norm induced by the ℓ^1-norm, we have

$$\begin{aligned}\mathbb{P}\left(||C^\varepsilon|| > t\right) &\leq n\mathbb{P}\Big(\sum_{j=1}^n |C^\varepsilon_{1,j}| > t\Big)\\ &= n\mathbb{P}\Big(\sum_{j=1}^n |X_{1,j}|\mathbb{1}_{\varepsilon n^{1/2}\leq |X_{1,j}|_\infty \leq \varepsilon^{-1}n^{1/2}} > t\sqrt{n}\Big).\end{aligned}$$

Since $|X_{i,j}|_\infty \leq |X_{i,j}| \leq \sqrt{2}|X_{i,j}|_\infty$,

$$\mathbb{P}\left(||C^\varepsilon|| > t\right) \leq n\mathbb{P}\Big(\sum_{j=1}^n Y_j > t\sqrt{n}\Big),$$

with $Y_j = |X_{1,j}|\mathbb{1}_{\varepsilon n^{1/2}\leq |X_{1,j}| \leq \sqrt{2}\varepsilon^{-1}n^{1/2}}$. But from Lemma 3.6.2 we deduce

$$\mathbb{E}Y_j = O\big(e^{-\frac{\kappa}{2}\varepsilon^\alpha n^{\alpha/2}}\big) = o\left(1/\sqrt{n}\right).$$

This yields for N large enough,

$$\mathbb{P}\Big(\sum_{j=1}^n Y_j > t\sqrt{n}\Big) \leq \mathbb{P}\Big(\sum_{j=1}^n (Y_j - \mathbb{E}Y_j) > \frac{t}{2}\sqrt{n}\Big). \tag{3.22}$$

But by Bennett's inequality (see [76, Theorem 2.9]), we have

$$\mathbb{P}\Big(\sum_{j=1}^n (Y_j - \mathbb{E}Y_j)\Big) > \frac{t}{2}\sqrt{n}\Big) \leq \exp\Big(-\frac{v}{2\varepsilon^{-2}n} h\Big(\frac{\varepsilon^{-1}nt}{\sqrt{2}v}\Big)\Big),$$

with $h(x) = (x+1)\log(x+1) - x$, and $v = \sum_{j=1}^n Y_j^2$. Using again Lemma 3.6.2, we find,

$$v = O\Big(ne^{-\frac{\kappa}{2}\varepsilon^\alpha n^{\alpha/2}}\Big). \tag{3.23}$$

As $h(x) \underset{x\to+\infty}{\sim} x\log x$, we have for n large enough,

$$\mathbb{P}\Big(\sum_{j=1}^n (Y_j - \mathbb{E}Y_j) > \frac{t}{2}\sqrt{n}\Big) \leq \exp\Big(-\frac{t}{2\sqrt{2}\varepsilon^{-1}} \log\Big(\frac{\varepsilon^{-1}nt}{\sqrt{2}v}\Big)\Big).$$

Using (3.23), we get

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\Big(\sum_{j=1}^n (Y_j - \mathbb{E}Y_j) > \frac{t}{2}\sqrt{n}\Big) \leq -\frac{\kappa}{4\sqrt{2}} t\varepsilon^{\alpha+1}. \tag{3.24}$$

Putting together inequalities (3.6) and (3.22) with the last exponential estimate (3.24), we get the claim

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(||C^\varepsilon|| > t\right) \leq -\frac{\kappa}{4\sqrt{2}} t\varepsilon^{\alpha+1}.$$

□

Finally, we now turn to the estimation of the last event $\{\lambda_{D^\varepsilon} > t\}$. It will directly fall from the following lemma.

3.6.6 Lemma. *For all $t > 0$,*

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(||D^\varepsilon|| > t\right) \leq -\frac{\kappa}{2}\varepsilon^{-\alpha}.$$

where κ is as in (3.9).

Proof. Just as in the proof of Lemma 3.6.5, we have

$$\mathbb{P}\left(||D^\varepsilon|| > t\right) \leq n\mathbb{P}\Big(\sum_{j=1}^{n} \frac{|X_{1,j}|}{\sqrt{n}} \mathbb{1}_{\varepsilon^{-1}n^{1/2}<|X_{1,j}|} > t\Big).$$

By Markov's inequality we get

$$\mathbb{P}\left(||D^\varepsilon|| > t\right) \leq \frac{\sqrt{n}}{t}\sum_{j=1}^{n} \mathbb{E}\big(|X_{1,j}|\mathbb{1}_{\varepsilon^{-1}n^{1/2}<|X_{1,j}|}\big).$$

From Lemma 3.6.2 we deduce

$$\mathbb{E}\big(|X_{1,j}|\mathbb{1}_{\varepsilon^{-1}n^{1/2}<|X_{1,j}|}\big) = O\big(e^{-\frac{\kappa}{2}\varepsilon^{-\alpha}n^{\alpha/2}}\big).$$

Therefore,

$$\mathbb{P}\left(||D^\varepsilon|| > t\right) = O\big(n\sqrt{n}e^{-\frac{\kappa}{2}\varepsilon^{-\alpha}n^{\alpha/2}}\big),$$

which gives the claim. □

Putting together the different estimates (3.15), (3.20), (3.21) and (3.6.6), and using inequality (3.10), we get

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}(\lambda_{X/\sqrt{n}} > 4t) \leq -C_1 \min\left(t^2\varepsilon^{-2+\alpha}, t\varepsilon^{\alpha+1}, \varepsilon^{-\alpha}\right), \tag{3.25}$$

where C_1 is some constant small enough. Taking the limsup as t goes to infinity, and then the limsup as ε goes to 0, we get finally

$$\limsup_{t\to+\infty}\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\lambda_{X/\sqrt{n}} > 4t\big) \leq -\infty.$$

□

We show now that at the exponential scale we consider, C^ε has a bounded number of non-zero entries. This will be crucial later when we will see C^ε as a finite rank perturbation of the matrix A.

3.6.7 Proposition. *For all $\varepsilon > 0$,*

$$\lim_{r\to+\infty} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(\mathrm{Card}\{(i,j) : C^{\varepsilon}_{i,j} \neq 0\} > r\right) = -\infty.$$

Proof. We follow here the argument of the proof of in [29, Lemma 2.2]. We have,

$$\mathbb{P}\left(\mathrm{Card}\{(i,j) : C^{\varepsilon}_{i,j} \neq 0\} > r\right) \leq \mathbb{P}\Big(\sum_{i\leq j} \mathbb{1}_{|X_{i,j}|\geq \varepsilon n^{1/2}} > r/2\Big).$$

Let $p_{i,j} = \mathbb{P}\left(|X_{i,j}| \geq \varepsilon n^{1/2}\right)$. From (3.9), we get that $p_{i,j} = o(n^{-2})$. Therefore it is enough to show that,

$$\limsup_{r\to+\infty} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\Big(\sum_{i\leq j}\Big(\mathbb{1}_{|X_{i,j}|\geq\varepsilon n^{1/2}} - p_{i,j}\Big) > r\Big) = -\infty.$$

Using Bennett's inequality (see [76, Theorem 2.9]), we get

$$\mathbb{P}\Big(\sum_{i\leq j}\big(\mathbb{1}_{|X_{i,j}|\geq\varepsilon n^{1/2}} - p_{i,j}\big) > r\Big) \leq \exp\Big(-vh\Big(\frac{r}{v}\Big)\Big),$$

with $h(x) = (x+1)\log(x+1) - x$, and $v = \sum_{i\leq j} p_{i,j}$. As $h(x) \underset{+\infty}{\sim} x\log x$, we have for n large enough,

$$\begin{aligned}\mathbb{P}\Big(\sum_{i\leq j}\big(\mathbb{1}_{|X_{i,j}|\geq\varepsilon n^{1/2}} - p_{i,j}\big) > r\Big) &\leq \exp\left(-r\log\left(\frac{r}{v}\right)\right)\\ &\leq \exp\left(r\log(rn^2)\right)\exp\big(-r\kappa\varepsilon^{\alpha}n^{\alpha/2}\big), \qquad (3.26)\end{aligned}$$

where we used in the last inequality the fact that $v \leq n^2 e^{-\kappa\varepsilon^{\alpha}n^{\alpha/2}}$, with κ as in (3.9). Taking the limsup at the exponential scale in (3.26), we get the claim. □

As a consequence of the latter proposition, we get the following result.

3.6.8 Proposition. *For all $\varepsilon > 0$,*

$$\lim_{r\to+\infty} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(\mathrm{rank}\left(C^{\varepsilon}\right) > r\right) = -\infty.$$

Proof. As the rank of a matrix is bounded by the number of non-zero entries, we see that Proposition 3.6.7 yields the claim. □

3.7 Exponential equivalences

3.7.1 First step

We show here that we can neglect at the exponential scale $n^{\alpha/2}$, the contributions of the very large entries (namely those such that $|X_{i,j}|_\infty > \varepsilon^{-1}\sqrt{n}$) and the intermediate entries (namely those such that $(\log n)^d < |X_{i,j}|_\infty < \varepsilon\sqrt{n}$) to the deviations of the largest eigenvalue of $X/\sqrt{n}$.

3.7.1 Proposition. *For all $t > 0$,*

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{A+C^\varepsilon} - \lambda_{X/\sqrt{n}}| > t\right) = -\infty,$$

where A and C^ε are as in (3.6). In short, $(\lambda_{A+C^\varepsilon})_{n \in \mathbb{N}, \varepsilon > 0}$ are exponentially good approximations of $(\lambda_{X/\sqrt{n}})_{n \in \mathbb{N}}$.

Proof. We have by Weyl's inequality (see (1.6)),

$$\mathbb{P}\left(|\lambda_{A+C^\varepsilon} - \lambda_{X/\sqrt{n}}| > t\right) \leq \mathbb{P}\left(||B^\varepsilon|| > t/2\right) + \mathbb{P}\left(||D^\varepsilon|| > t/2\right). \tag{3.27}$$

But we know by Lemma 3.6.6 and 3.6.4, that

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(||D^\varepsilon|| > \frac{t}{2}\big) \leq -\frac{\kappa}{2}\varepsilon^{-\alpha},$$

and

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\mathrm{tr}\,(B^\varepsilon)^2 > \frac{t}{2}\big) \leq -\frac{2^{\alpha/2}}{16} t\kappa\alpha\varepsilon^{-2+\alpha},$$

with κ as in (3.9). Thus, taking the limsup at the exponential scale $n^{\alpha/2}$ in (3.27), and then the limsup as ε goes to 0, recalling that $\alpha < 2$, we get the claim.

$\square$

3.7.2 Second step

We now show that in the study of the deviations of $\lambda_{A+C^\varepsilon}$, we can consider A and C^ε to be independent. We will prove the following result.

3.7.2 Theorem. *Let P_n denote the law of $X_{1,1}$ conditioned on the event $\{|X_{1,1}|_\infty \leq (\log n)^d\}$ and Q_n the law of $X_{1,2}$ conditioned on the event $\{|X_{1,2}|_\infty \leq (\log n)^d\}$. Let H be a random Hermitian matrix independent of X such that $(H_{i,j})_{1 \leq i \leq j \leq n}$ are independent, and for $1 \leq i \leq n$, $H_{i,i}$ has law P_n, and for all $i < j$, $H_{i,j}$ has law Q_n. We denote by H_n the normalized matrix $H/\sqrt{n}$.*

For all $t > 0$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \lambda_{H_n + C^\varepsilon}| > t\big) = -\infty.$$

With a similar argument as in the proof of Proposition 3.6.7, we get the following lemma.

3.7.3 Lemma. *Let $I = \{(i,j) : |X_{i,j}|_\infty > (\log n)^d\}$. For all $t > 0$,*

$$\lim_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}(|I| > tn^{\alpha/2}) = -\infty.$$

Proof of Theorem 3.7.2. Due to Proposition 3.7.1, it is enough to prove for any $\varepsilon > 0$ and any $t > 0$,

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{A+C^\varepsilon} - \lambda_{H_n + C^\varepsilon}| > t\right) = -\infty.$$

We will follow the same coupling argument to remove the dependency between A and C^ε, as in the proof of in [29, Proposition 2.1].

Let $I = \{(i,j) : |X_{i,j}|_\infty > (\log n)^d\}$. Let A' be the $n \times n$ matrix with (i,j)-entry,

$$A'_{i,j} = \mathbb{1}_{(i,j)\notin I} A_{i,j} + \mathbb{1}_{(i,j)\in I} \frac{H_{i,j}}{\sqrt{n}}.$$

Let $\mathcal{F}$ be the σ-algebra generated by the random variables $X_{i,j}$ such that $(i,j) \in I$. Then A' and H_n are independent of $\mathcal{F}$ and have the same law. By Weyl's inequality (see (1.6)),

$$\begin{aligned} |\lambda_{A+C^\varepsilon} - \lambda_{A'+C^\varepsilon}|^2 &\leq \operatorname{tr}(A - A')^2 \\ &= \frac{1}{n} \sum_{i,j} \left(\mathbb{1}_{(i,j)\in I} |H_{i,j}|^2\right) \\ &\leq |I| \frac{(\log n)^{2d}}{n}. \end{aligned} \tag{3.28}$$

Let $t > 0$. Define the event $F = \left\{|I| < t^2 n/(\log n)^{2d}\right\}$. Then, by Lemma 3.7.3 we have,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(F^c\right) = -\infty. \tag{3.29}$$

But according to (3.28),

$$\mathbb{1}_F \left|\lambda_{A+C^\varepsilon} - \lambda_{A'+C^\varepsilon}\right| \leq t. \tag{3.30}$$

Thus,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{A+C^\varepsilon} - \lambda_{A'+C^\varepsilon}| > t\right) = -\infty.$$

But C^ε is $\mathcal{F}$-measurable, and conditioned by $\mathcal{F}$, A' is a random Hermitian matrix with up-diagonal entries independent and bounded by $(\log n)^d/\sqrt{n}$. According to Proposition 3.5.1, we have

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{A'+C^\varepsilon} - \mathbb{E}_{\mathcal{F}}(\lambda_{A'+C^\varepsilon})| > t\right) = -\infty,$$

where $\mathbb{E}_{\mathcal{F}}$ denotes the conditional expectation given $\mathcal{F}$. Applying again Proposition 3.5.1 to H_n and C^ε, we get

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{H_n+C^\varepsilon} - \mathbb{E}_{\mathcal{F}}(\lambda_{H_n+C^\varepsilon})| > t\right) = -\infty.$$

But A' and H_n are independent of $\mathcal{F}$ and have the same law. Therefore,

$$\mathbb{E}_{\mathcal{F}}(\lambda_{A'+C^\varepsilon}) = \mathbb{E}_{\mathcal{F}}(\lambda_{H_n+C^\varepsilon}).$$

Thus by triangular inequality,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{A+C^\varepsilon} - \lambda_{H_n+C^\varepsilon}| > 3t\right) = -\infty,$$

which ends the proof. $\square$

3.8 Exponential approximation for the equation of the outliers

As a consequence of the LDP for the empirical spectral measure proved in [29], we show in the next proposition that the deviations at the left of 2 have an infinite cost at the exponential scale $n^{\alpha/2}$. This result will allow us to focus only on understanding the deviations of the largest eigenvalue at the right of 2.

3.8.1 Proposition.

$$\forall x < 2, \quad \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}(\lambda_{X/\sqrt{n}} \leq x) = -\infty.$$

Proof. Let $x < 2$. As the map $\mu \in \mathcal{P}(\mathbb{R}) \mapsto \sup \operatorname{supp}\mu \in \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous, we deduce that

$$F = \big\{\mu \in \mathcal{P}(\mathbb{R}) : \sup \operatorname{supp}\mu \leq x\big\},$$

is a closed set which does not contain μ_{sc}. As the spectral measure $\mu_{X/\sqrt{n}}$ satisfies a LDP with speed $n^{1+\alpha/2}$, and with good rate function I which achieves 0 only for the semi-circular law μ_{sc}, we have $\inf_F I_\alpha > 0$. Thus,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}(\lambda_{X/\sqrt{n}} \leq x) = -\infty.$$

$\square$

In the view of Theorem 3.7.2, Proposition 3.8.3, and Proposition 3.8.1, we are reduced to understand the deviations in $(2, +\infty)$, at the exponential scale $n^{\alpha/2}$, of the largest eigenvalue of the perturbed matrix $H_n + C^\varepsilon$, where C^ε can be assumed, due to Proposition 3.6.8 to be a finite rank matrix. We will use here the same approach as in many papers on finite rank deformations of Wigner matrices (see for example [20] or [78]) to determine the behavior of the extreme eigenvalues outside the bulk of a perturbed Wigner matrix. This approach is based on a determinant computation, stated here without proof, in the following lemma. It is a direct consequence of the Frobenius formula, which says that for any two matrices $A \in \mathcal{M}_{n,m}(\mathbb{C})$, $B \in \mathcal{M}_{m,n}(\mathbb{C})$, where $\mathcal{M}_{n,m}(\mathbb{C})$ denotes the set of matrices of size $n \times m$ with complex coefficients, we have

$$\det(I_n - AB) = \det(I_m - BA).$$

3.8.2 Lemma. *Let H and C be two Hermitian matrices of size N. Denote by k the rank of C, by $\theta_1, ..., \theta_k$ the non-zero eigenvalues of C in nondecreasing order and $u_1, ..., u_k$ orthonormal eigenvectors associated with these eigenvalues. Let $\sigma(H)$ be the spectrum of H. If $\lambda_{H+C} \notin \sigma(H)$, then it is the largest zero of f_n, where f_n is defined for all $z \notin \sigma(H)$ by*

$$f_n(z) = \det(M_n(z)), \text{ where } M_n(x) = I_k - \big(\theta_i \langle u_i, (x - H)^{-1} u_j\rangle\big)_{1\leq i,j\leq k}.$$

To make this strategy work, we need a control on the spectrum of H_n which will allow us to assume that the spectrum of H_n is nearly included $(-\infty, 2]$ at the exponential scale we consider. As a consequence of Proposition 3.5.1, and arguing similarly as in the proof of Corollary 3.6.3, we get the following proposition.

3.8.3 Proposition (Control on the spectrum of H_n)**.** *Let $\delta > 0$. Define*

$$C_\delta = \left\{ X \in \mathcal{H}_n^{(\beta)} : \lambda_X < 2 + \delta \right\}.$$

Then,

$$\lim_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(H_n \notin C_\delta \right) = -\infty,$$

with H_n is as in Theorem 3.7.2.

The goal of this section is to prove an exponential approximation for the equation of the eigenvalues of the perturbed matrix on every compact subset of $(2, +\infty)$. We will prove the following result.

3.8.4 Theorem. *Let H_n be as in Theorem 3.7.2 and let C_n be an independent random Hermitian matrix. Let k be the rank of C_n, $\theta_1, ..., \theta_k$ the non-zero eigenvalues in non-decreasing order of C_n and $u_1, ..., u_k$ orthonormal eigenvectors of C_n associated with these eigenvalues.*

Let $\delta > 0$, $\rho > 0$, and $r \in \mathbb{N}$. Define the event

$$W = \left\{ \mathrm{rank}(C_n) = r,\ ||C_n|| \leq \rho,\ \lambda_{H_n} \leq 2 + \delta \right\}. \tag{3.31}$$

For any $t > 0$, and any compact subset K of $(2 + \delta, +\infty)$,

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\Big(\Big\{ \sup_{x \in K} |f_n(x) - f(x)| > t \Big\} \cap W \Big) = -\infty,$$

where f_n is defined for any $x \notin Sp(H_n)$ by

$$f_n(x) = \det(M_n(x)),\ \text{with } M_n(x) = I_k - \left(\theta_i \langle u_i, (x - H_n)^{-1} u_j \rangle \right)_{1 \leq i,j \leq k},$$

and f, for any $x > 2$ by

$$f(x) = \det(M(x)),\ \text{with } M(x) = I_k - \left(\theta_i \delta_{i,j} g_{\mu_{sc}}(x) \right)_{1 \leq i,j \leq k}.$$

3.8.1 First step

We start by showing that M_n is close to its conditional expectation given C_n. As a consequence of Proposition 3.5.2, we get the following concentration result.

3.8.5 Proposition. *Define for all $x > 2 + \delta$,*

$$G^{(\delta)}(x) = \mathbb{1}_{H_n \in C_\delta} (x - H_n)^{-1},$$

where H_n is as in Theorem 3.7.2, and $C_\delta = \{X \in \mathcal{H}_n^{(\beta)} : \lambda_X < 2 + \delta\}$. Let u, v be two unit vectors. For any $t > 0$,

$$\lim_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \sup_{||u|| = ||v|| = 1} \mathbb{P}\Big(\big| \langle u, G^{(\delta)}(x) v \rangle - \mathbb{E}\langle u, G^{(\delta)}(x) v \rangle \big| > t \Big) = -\infty.$$

Proof. By the polarization formula we see that we only need to prove the claim for $u = v$. By assumption, H_n has its entries bounded by $(\log n)^d/\sqrt{n}$. Applying Proposition 3.5.2 with $\mu = 2+\delta$, we get that for any $x > 2+\delta$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \sup_{||u||=1} \mathbb{P}\left(\left|\tilde{f}_u(H_n) - \mathbb{E}(\tilde{f}_u(H_n))\right| > t\right) = -\infty, \tag{3.32}$$

where $\tilde{f}_u$ is a convex extension of f_u which is defined on C_δ by

$$f_u(Y) = \langle u, (x-Y)^{-1} u\rangle.$$

Furthermore, $\widetilde{f}_u$ is $1/(x-2-\delta)^2$-Lipschitz, with respect to the Hilbert-Schmidt norm. We have for all $t > 0$,

$$\mathbb{P}\left(\left|\tilde{f}_u(H_N) - \langle u, G^{(\delta)}(x)u\rangle\right| > t\right) \leq \mathbb{P}(\lambda_{H_n} \notin C_\delta), \tag{3.33}$$

which, invoking Proposition 3.8.3 yields,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \sup_{||u||=1} \mathbb{P}(|\tilde{f}_u(H_n) - \langle u, G^{(\delta)}(x)u\rangle| > t) = -\infty. \tag{3.34}$$

Moreover,

$$\left|\tilde{f}_u(H_n) - \langle u, G^{(\delta)}(x)u\rangle\right| \leq \mathbb{1}_{\lambda_{H_n}\notin C_\delta} \sup_{\mathcal{K}_n} \left|\tilde{f}_u\right|,$$

where the supremum is taken over the set $\mathcal{K}_n$ of Hermitian matrices of size n with entries bounded by $(\log n)^d/\sqrt{n}$. Thus,

$$\mathbb{E}\left|\tilde{f}_u(H_n) - \langle u, G^{(\delta)}(x)u\rangle\right| \leq \sup_{\mathcal{K}_n} \left|\tilde{f}_u\right| \mathbb{P}\left(\lambda_{H_n} \notin C_\delta\right). \tag{3.35}$$

It only remains to show that

$$\sup_{||u||=1} \mathbb{E}\left|\tilde{f}_u(H_n) - \langle u, G^{(\delta)}(x)u\rangle\right| \underset{N\to+\infty}{\longrightarrow} 0. \tag{3.36}$$

Indeed, putting together (3.32) with (3.34) and the claim above, we will get by the triangular inequality,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \sup_{||u||=1} \mathbb{P}\big(|\langle u, G^{(\delta)}(x)u\rangle - \mathbb{E}\langle u, G^{(\delta)}(x)u\rangle| > 2t\big) = -\infty.$$

We now show (3.36). Since $x > 2+\delta$, we have for all $H' \in C_\delta$,

$$\left|f_u(H')\right| \leq \frac{1}{\eta},$$

with $\eta = x - (2+\delta)$. Let H be a Hermitian matrix with entries bounded by $(\log n)^d/\sqrt{n}$. We have,

$$\left|\tilde{f}_u(H)\right| \leq \left|\tilde{f}_u(H) - \tilde{f}_u\Big(\frac{H}{||H||+1}\Big)\right| + \left|\tilde{f}_u\Big(\frac{H}{||H||+1}\Big)\right|.$$

But $H/(||H||+1)$ is in C_δ, thus $|f_u(H/(||H||+1))| \leq \frac{1}{\eta}$. Besides $\tilde{f}_u$ is $1/\eta^2$-Lipschitz with respect to the Hilbert-Schmidt norm. Therefore,

$$\begin{aligned}|\tilde{f}_u(H)| &\leq \frac{1}{\eta^2}||H||_2 + \frac{1}{\eta} \\ &\leq \frac{\sqrt{n}(\log n)^d}{\eta^2} + \frac{1}{\eta} \leq \frac{2\sqrt{n}(\log n)^d}{\eta^2}.\end{aligned}$$

We deduce that

$$\sup_{\mathcal{K}_n} |\tilde{f}_u| \leq \frac{2\sqrt{n}(\log n)^d}{\eta^2}.$$

From Proposition 3.8.3 we get,

$$\mathbb{E}|\tilde{f}_u(H_n) - \langle u, G^{(\delta)}(x)u\rangle| \underset{n\to+\infty}{\longrightarrow} 0,$$

which ends the proof of the claim. □

We are now ready to prove that M_n, restricted to the event that the spectrum of H_n is in $(-\infty, 2+\delta)$ for some $\delta > 0$, is exponentially equivalent to its conditional expectation given C_n, uniformly on any compact subset of $(2+\delta, +\infty)$.

3.8.6 Proposition (Concentration in the equation of eigenvalues outside the bulk)**.** *Let H_n be as in Theorem 3.7.2, and let C_n be an independent random Hermitian matrix. Let k be the rank of C_n, $\theta_1, ..., \theta_k$ the non-zero eigenvalues in non-decreasing order, and $u_1, ..., u_k$ orthonormal eigenvectors associated with these eigenvalues. For all $x > 2+\delta$, we define*

$$M_n^{(\delta)}(x) = I_k - \big(\theta_i\langle u_i, \mathbb{1}_{H_n\in C_\delta}(x-H_n)^{-1}u_j\rangle\big)_{1\leq i,j\leq k},$$

where $C_\delta = \{X \in \mathcal{H}_n^{(\beta)} : \lambda_X < 2+\delta\}$, and where H_n is as in Theorem 3.7.2.

Let $t > 0$ and $\rho > 0$. For any compact subset K of $(2+\delta, +\infty)$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{\sup_{x\in K} |M_n^{(\delta)}(x) - \mathbb{E}_{C_n}(M_n^{(\delta)}(x))|_\infty > t\} \cap V\big) = -\infty,$$

where

$$V = \{\mathrm{rank}(C_n) = r,\ ||C_n|| \leq \rho\},$$

and $\mathbb{E}_{C_n}$ denotes the conditional expectation given C_n, and where for any matrix M, $|M|_\infty = \sup_{i,j}|M_{i,j}|$.

Proof. Fix x in $(2+\delta, +\infty)$ and $i, j \in \{1, ..., r\}$. We will denote by $\mathbb{P}_{C_n}$ the conditional probability given C_n. We have,

$$\begin{aligned}&\mathbb{1}_V \mathbb{P}_{C_n}\big(\big|M_n^{(\delta)}(x)_{i,j} - \mathbb{E}_{C_n}(M_n^{(\delta)}(x)_{i,j})\big| > t\big) \\ &\qquad\leq \sup_{||u||=||v||=1} \mathbb{P}\big(\rho\big|\langle u, G^{(\delta)}(x)v\rangle - \mathbb{E}\langle u, G^{(\delta)}(x)v\rangle\big| > t\big),\end{aligned}$$

where $G^{(\delta)}(x)$ is as in Proposition 3.8.5. Thus, from Proposition 3.8.5, we get

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{|M_n^{(\delta)}(x)_{i,j} - \mathbb{E}_{C_n} M_n^{(\delta)}(x)_{i,j}| > t\} \cap V\big) = -\infty.$$

Taking the union over all the i, j in $\{1, ..., r\}$, we get for any $x \in (2+\delta, +\infty)$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{|M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)|_\infty > t\} \cap V\big) = -\infty.$$

We now use a chaining argument to extend this exponential equivalence uniformly in z in a given compact subset K of $(2+\delta, +\infty)$. Let $m \in \mathbb{N}$. Since K is compact, there are a finite number of points in $\{x \in K : mx \in \mathbb{Z}\}$. Taking the union bound, we deduce that for any $t > 0$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{\sup_{\substack{x\in K\\ mx\in\mathbb{Z}}} |M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)|_\infty > t\} \cap V\big) = -\infty. \tag{3.37}$$

Note that provided $\rho(C^\varepsilon) \le \rho$, we have for any $x, y \in K$,

$$\Big|M_n^{(\delta)}(x) - M_n^{(\delta)}(y)\Big|_\infty \le \frac{\rho}{\eta^2}|x-y|,$$

where $\eta = \inf K - (2+\delta)$. Therefore, on the event V, the function $x \in K \mapsto M_n^{(\delta)}(x)$ is ρ/η^2-Lipschitz with respect to the norm $|\ |_\infty$, and we have,

$$\sup_{x\in K} \big|M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)\big|_\infty \le \sup_{\substack{x\in K\\ mx\in\mathbb{Z}}} \big|M_m^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)\big|_\infty + \frac{2\rho}{m\eta^2}.$$

Taking m large enough, we get from (3.37) and the inequality above, that for any $t > 0$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{\sup_{x\in K} \big|M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)\big|_\infty > t\} \cap V\big) = -\infty.$$

□

3.8.2 Second step

The second step of the proof of Theorem 3.8.4 will be to prove an isotropic-like property of the semicircular law. This will be made possible due to the results on estimates of the coefficients of the resolvent of Wigner matrices in [81]. This is where our assumption on the independence between the real and imaginary parts of the entries of our Wigner matrix X plays its role.

3.8.7 Theorem. *For any compact subset K of $(2+\delta, +\infty)$,*

$$\sup_{x\in K} \sup_{||u||=||v||=1} |\langle u, \mathbb{E}G^{(\delta)}(x)v\rangle - \langle u, v\rangle g_{\mu_{sc}}(x)| \underset{n\to+\infty}{\longrightarrow} 0,$$

where $G^{(\delta)}$ is as in Proposition 3.8.5.

Proof. Let u and v be two unit vectors. Let K be a compact subset of $(2+\delta, +\infty)$. Set $\eta = \inf K - (2+\delta)$. To ease the notation, we denote for any $z \notin \sigma(H_n)$, the resolvent of H_n, $G(z) = (z - H_n)^{-1}$. Let $y > 0$ and $x \in K$. We write $z = x + iy$. We have,

$$\mathbb{1}_{H_n \in C_\delta} |\langle u, G(x)v\rangle - \langle u, G(z)v\rangle| \le \frac{y}{\eta^2}.$$

Thus,

$$\mathbb{E}|\langle u, G^{(\delta)}(x)v\rangle - \langle u, G(z)v\rangle| \leq \frac{y}{\eta^2} + \frac{1}{y}\mathbb{P}\left(H_n \notin C_\delta\right).$$

From now on we take $y = 1/\log n$. From Proposition 3.8.3, we get uniformly for x in K,

$$\sup_{||u||=||v||=1} \mathbb{E}\big|\langle u, G^{(\delta)}(x)v\rangle - \langle u, G(z)v\rangle\big| \underset{n\to+\infty}{\longrightarrow} 0. \tag{3.38}$$

Thus, we only need to show,

$$\sup_{||u||=||v||=1} \big|\mathbb{E}\langle u, G(z)v\rangle - \langle u, v\rangle g_{\mu_{sc}}(x)\big| \underset{n\to+\infty}{\longrightarrow} 0,$$

uniformly for $x \in K$.

Expanding the scalar product and using the exchangeability of the entries of H_n, we get

$$\begin{aligned}\langle u, \mathbb{E}G(z)v\rangle &= \sum_{1\leq i,j\leq n} \overline{u_i}\mathbb{E}G_{i,j}(z)v_j \\ &= \langle u, v\rangle\, \mathbb{E}G_{1,1}(z) + \sum_{i\neq j} \overline{u_i}v_j\mathbb{E}G_{1,2}(z) \\ &= \langle u, v\rangle\, \frac{1}{n}\mathbb{E}\mathrm{tr}G(z) + \sum_{i\neq j} \overline{u_i}v_j\mathbb{E}G_{1,2}(z).\end{aligned}$$

Since u and v are unit vectors,

$$\big|\langle u, \mathbb{E}G(z)v\rangle - \langle u, v\rangle\mathbb{E}\Big(\frac{1}{n}\mathrm{tr}G(z)\Big)\big| \leq n|\mathbb{E}G_{1,2}(z)|. \tag{3.39}$$

But since the entries of X have finite fifth moment and their real and imaginary parts are independent, we have according to [81, Proposition 3.1],

$$\mathbb{E}G_{1,2}(X/\sqrt{n})(z) = O\Big(\frac{P_9\left(1/\left|\Im(z)\right|\right)}{n^{3/2}}\Big), \tag{3.40}$$

uniformly for $z \in \mathbb{C}\setminus\mathbb{R}$, where we denote by $G(X/\sqrt{n})$ the resolvent of $X/\sqrt{n}$, and where P_9 is a polynomial of degree 9. But recall from the proof of Proposition 3.7.2 that H_n has the same law as the matrix A', where A' is the $n\times n$ matrix such that

$$A'_{i,j} = \frac{X_{i,j}}{\sqrt{n}}\mathbb{1}_{|X_{i,j}|_\infty \leq (\log n)^d} + \frac{H_{i,j}}{\sqrt{n}}\mathbb{1}_{|X_{i,j}|_\infty > (\log n)^d}.$$

Thus,

$$\mathbb{E}G_{1,2}(z) = \mathbb{E}G(A')_{1,2}(z), \tag{3.41}$$

where $G(A')$ denotes the resolvent of A'. Using the resolvent equation we get,

$$n\mathbb{E}|G(A')_{1,2}(z) - G(X/\sqrt{n})_{1,2}(z)| \leq n\left(\log n\right)^2\mathbb{E}||A' - X/\sqrt{n}||_2 \tag{3.42}$$

But it is easy to see that

$$\mathbb{E}||A' - X/\sqrt{n}||_2 = o\Big(\frac{1}{n\left(\log n\right)^2}\Big),$$

since we know from Lemma 3.6.2 that

$$\mathbb{E}\big(|X_{i,j}|\mathbb{1}_{|X_{i,j}|>(\log n)^d}\big) = O\big(e^{-\frac{\kappa}{2}(\log n)^{d\alpha}}\big),$$

with κ as in (3.9) and $d\alpha > 1$. Thus, the latter estimate, together with (3.42) and (3.41), yields,

$$n\big(\mathbb{E}G_{1,2}(z) - \mathbb{E}G(X/\sqrt{n})_{1,2}(z)\big) \underset{n\to+\infty}{\longrightarrow} 0,$$

uniformly in $x \in K$, where $z = x + i(\log n)^{-1}$. Using (3.40), we get

$$n\mathbb{E}G_{1,2}(x + i(\log n)^{-1}) \underset{n\to+\infty}{\longrightarrow} 0, \tag{3.43}$$

uniformly in $x \in K$. By the same coupling argument as above, one can show that

$$\mathbb{E}\Big(\frac{1}{n}\mathrm{tr}G(X/\sqrt{n})(z)\Big) - \mathbb{E}\Big(\frac{1}{n}\mathrm{tr}G(z)\Big) \underset{n\to+\infty}{\longrightarrow} 0,$$

uniformly for x in K.

But according to [81, Proposition 3.1], we have also

$$\mathbb{E}\Big(\frac{1}{n}\mathrm{tr}G(X/\sqrt{n})(z)\Big) = g_{\mu_{sc}}(z) + O\Big(\frac{1}{|\Im(z)|^6\, n}\Big),$$

uniformly on bounded subsets of $\mathbb{C}\setminus\mathbb{R}$. We deduce that,

$$\mathbb{E}\Big(\frac{1}{n}\mathrm{tr}G\Big(x + \frac{i}{\log n}\Big)\Big) \underset{n\to+\infty}{\longrightarrow} g_{\mu_{sc}}(x), \tag{3.44}$$

uniformly for x in K. Thus, putting (3.44) , (3.43) together with (3.39), we get

$$\sup_{||u||=||v||=1} |\langle u, \mathbb{E}G(x + i(\log n)^{-1})v\rangle - \langle u, v\rangle g_{\mu_{sc}}(x)| \underset{n\to+\infty}{\longrightarrow} 0,$$

uniformly for x in K, which completes the proof. □

As a consequence of Proposition 3.8.6, and the isotropic property of Proposition 3.8.7, with the control on the spectrum of H_n proved in Proposition 3.8.3, we get the following exponential equivalent for M_n.

3.8.8 Proposition. *Let H_n be as in Theorem 3.7.2 and C_n be a random Hermitian matrix independent of H_n. Let k be the rank of C_n, $\theta_1, \theta_2, ..., \theta_k$ the non-zero eigenvalues of C_n in non-decreasing order, and $u_1, u_2, ..., u_k$ orthonormal eigenvectors associated with these eigenvalues. We define for $x \notin \sigma(H_n)$,*

$$M_n(x) = I_k - \big(\theta_i\langle u_i, (x - H_n)^{-1}u_j\rangle\big)_{1\le i,j\le k},$$

and for all $x > 2$,

$$M(x) = I_k - \begin{pmatrix} \theta_1 g_{\mu_{sc}}(x) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \theta_k g_{\mu_{sc}}(x) \end{pmatrix}.$$

Let $\delta > 0$ and $\rho > 0$. For any compact subset K of $(2+\delta, +\infty)$ and $t > 0$, we have

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(\{ \sup_{x\in K} |M_n(x) - M(x)|_\infty > t\} \cap W\big) = -\infty,$$

with

$$W = \{\mathrm{rank}(C_n) = r, ||C_n|| \le \rho, \lambda_{H_n} \le 2+\delta\}.$$

Proof. By triangular inequality, we have

$$\begin{aligned}
&\mathbb{P}\big(\{ \sup_{x\in K} |M_n(x) - M(x)|_\infty > t\} \cap W\big) \\
&\le \mathbb{P}\big(\{ \sup_{x\in K} |M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)|_\infty > t/2\} \cap V\big) \\
&+ \mathbb{P}\big(\{ \sup_{x\in K} |\mathbb{E}_{C_n} M_n^{(\delta)}(x) - M(x)|_\infty > t/2\} \cap V\big),
\end{aligned}$$

with

$$V = \{\mathrm{rank}\,(C_n) = r, \rho(C_n) \le \rho\}.$$

From Theorem 3.8.7, we know that

$$\sup_{x\in K} \mathbb{1}_V |\mathbb{E}_{C_n} M_n^{(\delta)}(x) - M(x)|_\infty \xrightarrow[n\to+\infty]{L^\infty} 0,$$

where the convergence takes place in the space of essentially bounded functions. Thus, for n large enough,

$$\mathbb{P}\big(\{ \sup_{x\in K} |M_n(x) - M(x)|_\infty > t\} \cap W\big) \le \mathbb{P}\big(\{ \sup_{x\in K} |M_n^{(\delta)}(x) - \mathbb{E}_{C_n} M_n^{(\delta)}(x)|_\infty > t/2\} \cap V\big),$$

which, applying Proposition 3.8.6, ends the proof. □

We are now ready to give the proof of Theorem 3.8.4.

Proof of Theorem 3.8.4. Let K be compact subset of $(2+\delta, +\infty)$. Assuming W occurs, we see that for all x in K, the matrices $M_n(x)$ and $M(x)$ have their spectral radii bounded by

$$1 + \rho \max\left(1, \frac{1}{d(2+\delta, K)}\right),$$

where $d(2+\delta, K)$ is the distance of $2+\delta$ from K. Therefore $M(x)$ and $M_N(x)$ remain in a compact set of $\mathcal{H}_r^{(\beta)}$. As the determinant is uniformly continuous on compact sets of $\mathcal{H}_r^{(\beta)}$, Theorem 3.8.8 yields the claim.

□

3.9 An exponentially good approximation for the largest eigenvalue

We are interested here in finding simple exponentially good approximations of $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$, which will allow us to derive a large deviation principle for $\lambda_{X/\sqrt{n}}$. To this end, define for all $n \in \mathbb{N}$ and $\varepsilon > 0$,

$$\rho_{n,\varepsilon} = \begin{cases} g_{\mu_{sc}}^{-1}\left(1/\lambda_{C^\varepsilon}\right) & \text{if } \lambda_{C^\varepsilon} \ge 1, \\ 2 & \text{if } \lambda_{C^\varepsilon} < 1. \end{cases} \tag{3.45}$$

We will show in this section the following result.

3.9.1 Theorem. *For all* $t > 0$

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \rho_{n,\varepsilon}| > t\big) = -\infty.$$

In other words, $(\rho_{n,\varepsilon})_{n\in\mathbb{N},\varepsilon>0}$ *are exponentially good approximations of* $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ *at the exponential scale* $n^{\alpha/2}$.

Since we know from Theorem 3.7.2 that $(\lambda_{H_n+C^\varepsilon})_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$, we only need to prove Theorem 3.9.1 with $\lambda_{H_n+C^\varepsilon}$ instead of $\lambda_{X/\sqrt{n}}$. For sake of clarity, we will focus first on finding an exponential equivalent of $\lambda_{H_n+C_n}$ where C_n is a general random Hermitian matrix independent of H_n, and then we will apply our result to the matrix C^ε to get Theorem 3.9.1.

We know by Lemma 3.8.2, that provided $\lambda_{H_n+C_n}$ is outside the spectrum of H_n, it is the largest zero of f_n defined for all $z \notin \sigma(H_n)$ by

$$f_n(z) = \det\big(I_k - \big(\theta_i\langle u_i, (z - H_n)^{-1} u_j\rangle\big)_{1\le i,j\le k}\big),$$

with k the rank of C_n, $\theta_1, \theta_2, ..., \theta_k$ are the non-zero eigenvalues of C_n in non-decreasing order and $u_1, u_2, ..., u_k$ are orthonormal eigenvectors associated with those eigenvalues. But from Theorem 3.8.4, we know that this function is arbitrary close to a certain limit function f on every compact subset of $(2, +\infty)$ with an overwhelming probability, with f defined for all $x \notin (-2, 2)$ by

$$f(x) = \prod_{i=1}^{k} (1 - \theta_i g_{\mu_{sc}}(x)) . \tag{3.46}$$

Therefore, one can hope that the largest zero of f_n, which is the top eigenvalue of $H_n + C_n$, is arbitrary close to the largest zero of f. From the equation verified by the Stieltjes transform of $g_{\mu_{sc}}$,

$$\forall z \in \mathbb{C} \setminus (-2, 2),\ g_{\mu_{sc}}(z) + \frac{1}{g_{\mu_{sc}}(z)} = z,$$

(see (1.5)) we see that $g_{\mu_{sc}}$ is decreasing on $[2, +\infty)$ taking its values in $(0, 1]$, and that

$$\forall x \in (0, 1],\ g^{-1}_{\mu_{sc}}(x) = x + \frac{1}{x}.$$

Thus, f admits a zero only when $\theta_k > 1$, in which case its largest zero is $g^{-1}_{\mu_{sc}}(1/\theta_k)$, that is, $g^{-1}_{\mu_{sc}}(1/\lambda_{C_n})$.

3.9.2 Proposition. *Let* H_n *be as in Theorem 3.7.2, and let* C_n *be a random Hermitian matrix independent of* H_n. *Let* $\delta > 0$ *and* $l \geq 2 + 2\delta$. *For all* $t > 0$ *and* $r \in \mathbb{N}$,

$$\lim_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(|\lambda_{H_n+C_n} - \rho_n| > t, \rho_n \geq 2 + 2\delta, \lambda_{H_n+C_n} \leq l, C_n \in V_{r,l}\right) = -\infty,$$

where

$$\rho_n = \begin{cases} g^{-1}_{\mu_{sc}}(1/\lambda_{C_n}) & \text{if } \lambda_{C_n} \geq 1, \\ 2 & \text{if } \lambda_{C_n} < 1. \end{cases}$$

and

$$V_{r,l} = \left\{C \in \mathcal{H}_n^{(\beta)} : \mathrm{rank}(C) = r, ||C|| \leq 1/g_{\mu_{sc}}(l)\right\}.$$

Proof. We start by reducing the problem to the case where C_n has its top eigenvalue simple and bounded away from its last-but-one eigenvalue. Let u be an eigenvector associated with the largest eigenvalue of C_n. Let $\gamma > 0$. We denote by $C_n^{(\gamma)}$ the matrix defined by,

$$C_n^{(\gamma)} = C_n + \gamma u u^*.$$

By definition, the largest eigenvalue of C_n is bounded away from its last-but-one eigenvalue by γ. Provided that $\lambda_{C_n} \geq 1$, we define

$$\mu_n^{(\gamma)} = g_{\mu_{sc}}^{-1}\left(1/\lambda_{C_{n,\gamma}}\right) = g_{\mu_{sc}}^{-1}\left(1/(\lambda_{C_n} + \gamma)\right).$$

Weyl's inequality (1.6) yields,

$$|\lambda_{H_n + C_n^{(\gamma)}} - \lambda_{H_n + C_n}| \leq \gamma.$$

As for all $x \in (0,1]$, $g_{\mu_{sc}}^{-1}(x) = x + \frac{1}{x}$, easy computation yields

$$|\mu_n^{(\gamma)} - \mu_n| \leq 2\gamma.$$

Thus, we see that it is sufficient to prove the statement in Proposition 3.9.2 but with $V_{r,l}^{(\gamma)}$ instead of $V_{r,l}$, where

$$V_{r,l}^{(\gamma)} = \left\{ C \in \mathcal{H}_n^{(\beta)} : \operatorname{rank}(C) = r, \ ||C|| \leq 1/g_{\mu_{sc}}(l), \ \theta_r(C) - \theta_{r-1}(C) \geq \gamma \right\},$$

where $\theta_r(C)$, and $\theta_{r-1}(C)$ denote respectively the largest and the second largest eigenvalue of C.

We know from Theorem 3.8.4 that the functions f_n and f are arbitrary close on any compact subset of $(2, +\infty)$, with exponentially high probability. Since we cannot make the error on the distance between f_n and f in Theorem 3.8.4 depends on C_n, we need now a kind of uniform continuity property of the largest zero of continuous functions belonging to a certain compact set, to get that their largest zeros are close with an overwhelming probability. This is the subject of the following lemma.

3.9.3 Lemma. *Let $K' \subset K$ be two compact subsets of $\mathbb{R}$, such that there is some open set U such that $K' \subset U \subset K$. Let $\mathcal{K}$ a compact subset of $C(K)$, the space of continuous functions on K taking real values. We assume that any $f \in \mathcal{K}$ admits at least one zero in K, its largest zero, z(f), lies in K', and f changes sign at $z(f)$. Then, for all $t > 0$, there is some $s > 0$, such that for all $f \in \mathcal{K}$ and $g \in C(K)$, such that*

$$||f - g|| < t,$$

g admits at least one zero in K, and its largest zero $z(g)$, satisfies

$$|z(f) - z(g)| < s.$$

Proof. As an consequence of the intermediate values theorem, the function φ, defined for all $g \in C(K)$ by,

$$\varphi(g) = \begin{cases} z_{\max}(g) & \text{if } g \text{ admits a zero in } K, \\ \dagger & \text{otherwise,} \end{cases}$$

is continuous at each $f \in \mathcal{K}$. As the set $\mathcal{K}$ is compact, we get the claim.

$\square$

We come back now at the proof of Proposition 3.9.2. Observe that if $C_n \in V_{r,l}^{(\gamma)}$, then $\mu_n \leq l$. Let K be a compact set such that there is an open set U satisfying $[2+2\delta, l] \subset U \subset K \subset (2+\delta, +\infty)$. Note that the subset

$$\mathcal{K}^{(\gamma)} = \big\{ x \in K \mapsto \prod_{i=1}^{r} (1 - \theta_i g_{\mu_{sc}}(x)) : (\theta_1, ..., \theta_r) \in \Theta_\gamma \big\},$$

where

$$\Theta^{(\gamma)} = \big\{ (\theta_1, ..., \theta_r) \in \mathbb{R}^r : -\rho \leq \theta_1 \leq ... \leq \theta_{r-1} \leq \theta_r - \gamma, \ 1 \leq \theta_r \leq \rho, \ g_{\mu_{sc}}^{-1}(1/\theta_r) \in K' \big\},$$

is a compact subset of $C(K)$. Applying Lemma 3.9.3 with $K' = [2+2\delta, l]$ and K, we get for any $t > 0$, that there is $s > 0$, such that

$$\begin{aligned}
&\mathbb{P}\big(|\lambda_{H_n + C_n} - \mu_n| > t, \mu_n \in K', \lambda_{H_n+C_n} \leq l, C_n \in V_{r,l}^{(\gamma)} \big) \\
&\quad \leq \mathbb{P}\big(\big\{ \sup_{x \in K} |f_n(x) - f(x)| > s \big\} \cap W \big) + \mathbb{P}\big(\lambda_{H_n} > 2 + \delta \big),
\end{aligned}$$

with

$$W = \big\{ \mathrm{rank}(C_n) = r, \ ||C_n|| \leq 1/g_{\mu_{sc}}(l), \ \lambda_{H_n} \leq 2 + \delta \big\}.$$

By Theorem 3.8.4 and Proposition 3.8.3, we deduce that,

$$\lim_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{H_n+C_n} - \rho_n| > t, \rho_n \geq 2 + 2\delta, \lambda_{H_n+C_n} \leq l, C_n \in V_{r,l}^{(\gamma)} \big) = -\infty,$$

which ends the proof of Proposition 3.9.2. $\square$

We are now ready to give the proof of Theorem 3.9.1.

Proof of Theorem 3.9.1 . According to Proposition 3.8.1, we only need to prove that for $\delta > 0$ small enough,

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \rho_{n,\varepsilon}| > t, \lambda_{X/\sqrt{n}} > 2 - \delta \big) = -\infty.$$

Taking $\delta < t/3$, we see that it is actually sufficient to show

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2 + 2\delta \big) = -\infty. \tag{3.47}$$

Using Proposition 3.9.2, but with C^ε instead of C_n, we get for any $l \geq 2 + 2\delta$, and $s \in \mathbb{N}$,

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{H_n + C^\varepsilon} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2 + 2\delta, \lambda_{H_n+C^\varepsilon} \leq l, C^\varepsilon \in V_{s,l} \big) = -\infty,$$

where $\rho_{n,\varepsilon}$ is defined as in (3.45), and where $V_{s,l}$ is defined in Proposition 3.9.2. Let $V_{\leq r,l} = \cup_{s=0}^{r} V_{s,l}$. Since $V_{\leq r,l}$ is a finite union of the $V_{s,l}$'s, we get

$$\limsup_{n \to +\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{H_n + C^\varepsilon} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2 + 2\delta, \lambda_{H_n+C^\varepsilon} \leq l, C^\varepsilon \in V_{\leq r,l} \big) = -\infty.$$

As a consequence of Lemma 3.6.5 and Proposition 3.6.8, we deduce that for any $\varepsilon > 0$,

$$\lim_{r,l\to+\infty} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\left(C^{\varepsilon} \notin V_{\leq r,l}\right) = -\infty.$$

Thus,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{H_n+C^{\varepsilon}} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2+2\delta, \lambda_{H_n+C^{\varepsilon}} \leq l\big) = -\infty.$$

Using the fact that according to Theorem 3.7.2, $(\lambda_{H_n+C^{\varepsilon}})_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\lambda_{X/\sqrt{n}})_{\in\mathbb{N}}$, we get,

$$\lim_{\varepsilon\to 0} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2+2\delta, \lambda_{X/\sqrt{n}} \leq l\big) = -\infty.$$

But $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ is exponentially tight according to Proposition 3.6.1, thus we can conclude that,

$$\lim_{\varepsilon\to 0} \limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \mathbb{P}\big(|\lambda_{X/\sqrt{n}} - \rho_{n,\varepsilon}| > t, \rho_{n,\varepsilon} \geq 2+2\delta\big) = -\infty,$$

which ends the proof. □

3.10 Large deviations principle for $\lambda_{X/\sqrt{n}}$

Our aim here is to prove for each $\varepsilon > 0$, a LDP for $(\rho_{\varepsilon,n})_{n\in\mathbb{N}}$. Since $(\rho_{n,\varepsilon})_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of the largest eigenvalue of $X/\sqrt{n}$, we will get a large deviations principle for $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$.

For every $r\in\mathbb{N}$, we define

$$\mathcal{E}_r = \{A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)} : \mathrm{Card}\{(i,j) : A_{i,j} \neq 0\} \leq r\}.$$

For any $m\in\mathbb{N}$, let $\mathfrak{S}_m$ be the symmetric group on the set $\{1,...,m\}$. We denote by $\mathfrak{S}$, the group $\cup_{n\in\mathbb{N}}\mathfrak{S}_n$. We denote by $\widetilde{\mathcal{E}}_r$ the set of equivalence classes of $\mathcal{E}_r$ under the action of $\mathfrak{S}$, which is defined by

$$\forall \sigma \in \mathfrak{S}, \forall A \in \mathcal{E}_r,\ \sigma.A = M_{\sigma}^{-1} A M_{\sigma} = \left(A_{\sigma(i),\sigma(j)}\right)_{i,j},$$

where M_{σ} denotes the permutation matrix associated with the permutation σ, that is, $M_{\sigma} = (\delta_{i,\sigma(j)})_{i,j}$.

Let $\mathcal{H}_r^{(\beta)}/\mathfrak{S}_r$ be the set of equivalence classes of $\mathcal{H}_n^{(\beta)}$ under the action of the symmetric group $\mathfrak{S}_r$. Note that any equivalence class of the action of $\mathfrak{S}$ on $\mathcal{E}_r$ has a representative in $\mathcal{H}_n^{(\beta)}$. This defines an injective map from $\widetilde{\mathcal{E}}_r$ into $\mathcal{H}_n^{(\beta)}/\mathfrak{S}_r$. Identifying $\widetilde{\mathcal{E}}_r$ to a subset of $\mathcal{H}_n^{(\beta)}/\mathfrak{S}_r$, we equip $\widetilde{\mathcal{E}}_r$ of the quotient topology of $\mathcal{H}_n^{(\beta)}/\mathfrak{S}_r$. This topology is metrizable by the distance $\tilde{d}$ given by

$$\forall \tilde{A}, \tilde{B} \in \widetilde{\mathcal{E}}_r,\ \tilde{d}\left(\tilde{A}, \tilde{B}\right) = \min_{\sigma,\sigma'\in\mathcal{S}} \max_{i,j} \left|B_{\sigma(i),\sigma(j)} - A_{\sigma'(i),\sigma'(j)}\right|, \tag{3.48}$$

where A and B are two representatives of $\tilde{A}$ and $\tilde{B}$ respectively. Since the application which associates to a matrix of $\mathcal{H}_n^{(\beta)}$ its largest eigenvalue is continuous and is invariant by conjugation, we can define this application on $\mathcal{H}_n^{(\beta)}/\mathfrak{S}_r$ and it will still be continuous. Therefore, the application which associates to a matrix of $\widetilde{\mathcal{E}}_r$ its largest eigenvalue is continuous for the topology we defined above. This fact will be crucial later when we will apply a contraction principle to derive a large deviations principle for $(\rho_{n,\varepsilon})_{n\in\mathbb{N},\varepsilon>0}$.

Let $\varepsilon > 0$. Let $\mathbb{P}^{\varepsilon}_{n,r}$ be the law of C^{ε}, with C^{ε} as in (3.6), conditioned on the event $\{C^{\varepsilon} \in \mathcal{E}_r\}$, and $\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}$ the push forward of $\mathbb{P}^{\varepsilon}_{n,r}$ by the projection $\pi : \mathcal{E}_r \to \widetilde{\mathcal{E}}_r$.

3.10.1 Proposition. *Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Then $(\widetilde{\mathbb{P}}^{\varepsilon}_{n,r})_{n\in\mathbb{N}}$ satisfies a LDP with speed $n^{\alpha/2}$, and good rate function $I_{\varepsilon,r}$ defined for all $\tilde{A} \in \widetilde{\mathcal{E}}_r$ by,*

$$I_{\varepsilon,r}\left(\tilde{A}\right) = \begin{cases} b\sum_{i\geq 1}|A_{i,i}|^{\alpha} + \frac{a}{2}\sum_{i\neq j}|A_{i,j}|^{\alpha} & \text{if } A \in \mathcal{D}_{\varepsilon,r}, \\ +\infty & \text{otherwise,} \end{cases} \tag{3.49}$$

where A is a representative of the equivalence class $\tilde{A}$ and

$$\mathcal{D}_{\varepsilon,r} = \left\{A \in \mathcal{E}_r : \forall i \leq j,\ A_{i,j} = 0 \text{ or } \varepsilon \leq |A_{i,j}| \leq \varepsilon^{-1}, \text{ and } A_{i,j}/|A_{i,j}| \in \operatorname{supp}(\nu_{i,j})\right\},$$

with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in (3.2.1).

Recall from remark 3.2.3, that the assumptions 3.2.1 we made on the tail distribution of the entries, and the independence of the angle and the modulus of the entries, yield the following deviations lower bounds.

3.10.2 Lemma. *For all $\delta > 0$, and all $x \neq 0$ with $x/|x| \in \operatorname{supp}(\nu_1)$, there is a sequence $(b_n)_{n\in\mathbb{N}}$ which converges to b, such that for n large enough,*

$$\mathbb{P}\big(X_{1,1}/\sqrt{n} \in B(x,\delta)\big) \geq e^{-b_n|x|^{\alpha}n^{\alpha/2}}.$$

Similarly, for all $z \neq 0$ such that $z/|z| \in \operatorname{supp}(\nu_2)$, and all $0 < \delta < |z|$, there is a sequence $(a_n)_{n\in\mathbb{N}}$ which converges to a, such that for n large enough,

$$\mathbb{P}\left(X_{1,2}/\sqrt{n} \in B(z,\delta)\right) \geq e^{-a_n|z|^{\alpha}n^{\alpha/2}}.$$

Proof of Proposition 3.10.1. **Property of the rate function:** The function W_{α} defined on $\mathcal{H}_r^{(\beta)}$ by,

$$W_{\alpha}(A) = b\sum_{i=1}^{r}|A_{i,i}|^{\alpha} + \frac{a}{2}\sum_{1\leq i\neq j\leq r}|A_{i,j}|^{\alpha},$$

has compact level sets. Thus, we can deduce, by definition of the topology we equipped $\widetilde{\mathcal{E}}_r$, that the rate function $I_{\varepsilon,r}$ has also compact level sets.

Exponential tightness: Let $\gamma > 0$. We define,

$$K_{\gamma} = \Big\{\tilde{A} \in \widetilde{\mathcal{E}}_r : \sum_{i,j\in\mathbb{N}}|A_{i,j}|^{\alpha} \leq \gamma\Big\},$$

where A denotes a representative of $\tilde{A}$. Since the set

$$\Big\{A \in \mathcal{H}_n^{(\beta)} : \sum_{1\leq i,j\leq r}|A_{i,j}|^{\alpha} \leq \gamma\Big\},$$

is a compact subset of $\mathcal{H}_n^{(\beta)}$ and invariant under the action of $\mathfrak{S}_r$, we can deduce, by the choice of the topology we equipped $\widetilde{\mathcal{E}}_r$, that $\widetilde{K}_\gamma$ is a compact subset of $\widetilde{\mathcal{E}}_r$. Then, by definition of $\widetilde{\mathbb{P}}^\varepsilon_{n,r}$, we have

$$\widetilde{\mathbb{P}}^\varepsilon_{n,r}\left(K_\gamma^c\right) = \mathbb{P}\Big(\sum_{1\le i,j\le n} \left|C^\varepsilon_{i,j}\right|^\alpha > \gamma \mid C^\varepsilon \in \mathcal{E}_r\Big). \tag{3.50}$$

But $\mathbb{1}_{\sum_{i,j}|C^\varepsilon_{i,j}|>\gamma}$ and $\mathbb{1}_{C^\varepsilon\in\mathcal{E}_r}$ are respectively nondecreasing and nonincreasing with respect to the absolute value of each entry of C^ε. Therefore, Harris' inequality yields,

$$\begin{aligned}\mathbb{P}\Big(\sum_{1\le i,j\le n} \left|C^\varepsilon_{i,j}\right|^\alpha > \gamma \mid C^\varepsilon \in \mathcal{E}_r\Big) &\le \mathbb{P}\Big(\sum_{1\le i,j\le n} \left|C^\varepsilon_{i,j}\right|^\alpha > \gamma \Big)\\ &\le \mathbb{P}\Big(\sum_{i=1}^n |C^\varepsilon_{i,j}|^\alpha > \gamma/2\Big) + \mathbb{P}\Big(\sum_{1\le i\ne j\le n} |C^\varepsilon_{i,j}|^\alpha > \gamma/2\Big).\end{aligned}$$

Now choose a_1 such that $0 < 2a_1 < a$, and b_1 such that $0 < b_1 < b$. By Chernoff's inequality we have,

$$\begin{aligned}\widetilde{\mathbb{P}}^\varepsilon_{n,r}\left(K_\gamma^c\right) &\le e^{-b_1 n^{\alpha/2}\frac{\gamma}{2}}\mathbb{E}\left(e^{b_1|X_{1,1}|^\alpha \mathbb{1}_{\varepsilon n^{1/2}\le|X_{1,1}|\le\varepsilon^{-1}n^{1/2}}}\right)^n\\ &\quad + e^{-a_1 n^{\alpha/2}\frac{\gamma}{2}}\mathbb{E}\left(e^{2a_1|X_{1,2}|^\alpha \mathbb{1}_{\varepsilon n^{1/2}\le|X_{1,2}|_\infty\le\varepsilon^{-1}n^{1/2}}}\right)^{n(n-1)/2}.\end{aligned} \tag{3.51}$$

Let $b_2 \in (b_1, b)$. For t large enough we have,

$$\mathbb{P}\left(|X_{1,1}| > t\right) \le e^{-b_2 t^\alpha}.$$

Thus, integrating by part just as in the proof of Lemma 3.6.4 we get, for n large enough,

$$\mathbb{E}\Big(e^{b_1|X_{1,1}|^\alpha \mathbb{1}_{\varepsilon n^{1/2}\le|X_{1,1}|\le\varepsilon^{-1}n^{1/2}}}\Big) \le \exp\Big(\frac{b_2}{b_2-b_1}e^{-(b_2-b_1)\varepsilon^\alpha n^{\alpha/2}}\Big). \tag{3.52}$$

Similarly, for n large enough and with a_2 such that $2a_2 \in (2a_1, a)$ we have,

$$\mathbb{E}\Big(e^{2a_1|X_{1,2}|^\alpha \mathbb{1}_{\varepsilon n^{1/2}\le|X_{1,2}|_\infty\le\varepsilon^{-1}n^{1/2}}}\Big) \le \exp\Big(\frac{a_2}{a_2-a_1}e^{-2(a_2-a_1)\varepsilon^\alpha n^{\alpha/2}}\Big). \tag{3.53}$$

Therefore, putting together (3.52) and (3.53) into (3.51), we get,

$$\limsup_{n\to+\infty} \frac{1}{n^{\alpha/2}} \log \widetilde{\mathbb{P}}^\varepsilon_{n,r}\left(\widetilde{K}_\gamma^c\right) \le -\frac{\gamma}{2} a_1 \vee b_1,$$

which proves that $(\widetilde{\mathbb{P}}^\varepsilon_{n,r})_{n\in\mathbb{N}}$ is exponentially tight.

Lower bound: Let $A \in \mathcal{H}_r^{(\beta)}$. Without loss of generality, we can assume that $I_{\varepsilon,r}(\tilde{A}) < +\infty$, that is $A \in \mathcal{D}_{\varepsilon,r}$. Moreover, we assume that for all $1 \le i,j \le r$,

$$A_{i,j} = 0 \text{ or } \varepsilon < |A_{i,j}| < \varepsilon^{-1}.$$

Let $\delta > 0$ be such that

$$\delta < \min\Big(\min_{A_{i,j}\ne 0} |A_{i,j}| - \varepsilon, \varepsilon^{-1} - \max_{1\le i,j\le r} |A_{i,j}|, \varepsilon\Big).$$

Let

$$\tilde{B}\left(\tilde{A},\delta\right)=\left\{\tilde{X}\in\widetilde{\mathcal{E}}_r:\tilde{d}\left(\tilde{A},\tilde{X}\right)<\delta\right\},$$

with $\tilde{d}$ being the distance defined in (3.48). We have

$$\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}\big(\tilde{B}\left(\tilde{A},\delta\right)\big)=\mathbb{P}\Big(\min_{\sigma\in\mathfrak{S}}\max_{i,j}|C^{\varepsilon}_{\sigma(i),\sigma(j)}-A_{i,j}|<\delta\mid C^{\varepsilon}\in\mathcal{E}_r\Big).$$

Let

$$B_{\infty,n}\left(A,\delta\right)=\big\{X\in\mathcal{H}^{(\beta)}_n:\max_{1\leq i,j\leq n}|X_{i,j}-A_{i,j}|<\delta\big\}.$$

Since $\delta<\varepsilon$, and since all the non-zero entries of C^{ε} are in $\{z\in\mathbb{C}:\varepsilon\leq|z|\leq\varepsilon^{-1}\}$, we see that if $C^{\varepsilon}\in B_{\infty,n}(A,\delta)$, then $C^{\varepsilon}\in\mathcal{E}_r$. Thus,

$$\begin{aligned}\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}\big(\tilde{B}(\tilde{A},\delta)\big)&\geq\mathbb{P}\left(C^{\varepsilon}\in B_{\infty,n}\left(A,\delta\right)|C^{\varepsilon}\in\mathcal{E}_r\right)\\&=\frac{1}{\mathbb{P}\left(C^{\varepsilon}\in\mathcal{E}_r\right)}\mathbb{P}\left(C^{\varepsilon}\in B_{\infty,n}\left(A,\delta\right)\right).\end{aligned}\tag{3.54}$$

But by independence, we have

$$\mathbb{P}\left(C^{\varepsilon}\in B_{\infty,n}\left(A,\delta\right)\right)=\prod_{i=1}^{n}\mathbb{P}\left(|C^{\varepsilon}_{i,i}-A_{i,i}|<\delta\right)\prod_{i<j}\mathbb{P}\left(|C^{\varepsilon}_{i,j}-A_{i,j}|<\delta\right).\tag{3.55}$$

Since

$$\delta<\min\Big(\min_{A_{i,j}\neq0}|A_{i,j}|-\varepsilon,\varepsilon^{-1}-\max_{1\leq i,j\leq r}|A_{i,j}|\Big),$$

we have

$$\mathbb{P}\left(|C^{\varepsilon}_{i,i}-A_{i,i}|<\delta\right)\geq\mathbb{P}\Big(\big|X_{i,i}/\sqrt{n}-A_{i,i}\big|<\delta\Big)\mathbb{1}_{A_{i,i}\neq0}+\mathbb{P}\left(C^{\varepsilon}_{i,i}=0\right)\mathbb{1}_{A_{i,i}=0}.$$

Thus, according to Lemma 3.10.2, there is a sequence $(b_n)_{n\in\mathbb{N}}$ converging to b such that,

$$\begin{aligned}\mathbb{P}\left(|C^{\varepsilon}_{i,i}-A_{i,i}|<\delta\right)&\geq e^{-b_n|A_{i,i}|^{\alpha}n^{\alpha/2}}\mathbb{1}_{A_{i,i}\neq0}+\big(1-\mathbb{P}(|C^{\varepsilon}_{i,i}|\neq0)\big)\mathbb{1}_{A_{i,i}=0}\\&\geq e^{-b_n|A_{i,i}|^{\alpha}n^{\alpha/2}}\mathbb{1}_{A_{i,i}\neq0}+\big(1-\mathbb{P}(|X_{i,i}|\geq\varepsilon n^{1/2})\big)\mathbb{1}_{A_{i,i}=0}.\end{aligned}$$

For n large enough we get, with κ defined in (3.5), we get

$$\begin{aligned}\mathbb{P}\left(|C^{\varepsilon}_{i,i}-A_{i,i}|<\delta\right)&\geq e^{-b_n|A_{i,i}|^{\alpha}n^{\alpha/2}}\mathbb{1}_{A_{i,i}\neq0}+\left(1-e^{-\kappa\varepsilon^{\alpha}n^{\alpha/2}}\right)\mathbb{1}_{A_{i,i}=0}\\&\geq e^{-b_n|A_{i,i}|^{\alpha}n^{\alpha/2}}\left(1-e^{-\kappa\varepsilon^{\alpha}n^{\alpha/2}}\right).\end{aligned}\tag{3.56}$$

Similarly for $i\neq j$, we have,

$$\mathbb{P}\left(|C^{\varepsilon}_{i,j}-A_{i,j}|<\delta\right)\geq e^{-a_n|A_{i,j}|^{\alpha}n^{\alpha/2}}\left(1-e^{-\kappa\varepsilon^{\alpha}n^{\alpha/2}}\right),\tag{3.57}$$

where $(a_n)_{n\in\mathbb{N}}$ is a sequence converging to a. Putting (3.56) and (3.57) into (3.55), we get,

$$\mathbb{P}\left(C^{\varepsilon}\in B_{\infty,n}\left(A,\delta\right)\right)\geq e^{-b_n\sum_{i\geq1}|A_{i,i}|^{\alpha}n^{\alpha/2}}e^{-a_n\sum_{i<j}|A_{i,j}|^{\alpha}n^{\alpha/2}}\left(1-e^{-\kappa\varepsilon^{\alpha}n^{\alpha/2}}\right)^{n^2}.$$

Hence at the exponential scale,

$$\liminf_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\left(C^{\varepsilon}\in B_{\infty,n}\left(A,\delta\right)\right)\geq -b\sum_{i\geq 1}|A_{i,i}|^{\alpha}-a\sum_{i<j}|A_{i,j}|^{\alpha}.$$

Besides by Proposition 3.6.7 and Borel-Cantelli Lemma, we have

$$\mathbb{P}\left(C^{\varepsilon}\in\mathcal{E}_r\right)\underset{n\to+\infty}{\longrightarrow}1.$$

Putting these estimates into (3.54), we get

$$\liminf_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}\left(\tilde{B}\left(\tilde{A},\delta\right)\right)\geq -b\sum_{i=1}^{r}|A_{i,i}|^{\alpha}-a\sum_{1\leq i<j\leq r}|A_{i,j}|^{\alpha}. \tag{3.58}$$

Observe that since the rate function $I_{\varepsilon,r}$ is continuous on its domain $\pi\left(\mathcal{D}_{\varepsilon,r}\right)$, we have also the bound (3.58) for any $A\in\mathcal{D}_{\varepsilon,r}$. This concludes the proof of the lower bound.

Upper bound: From our assumption 3.2.1, we deduce that for n large enough, the support of $\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}$ is included in the domain of $I_{\varepsilon,r}$, that is $\pi\left(\mathcal{D}_{\varepsilon,r}\right)$. Thus, we see that whenever $I_{\varepsilon,r}(\tilde{A})=+\infty$ for $\tilde{A}\in\widetilde{\mathcal{E}}_r$,

$$\lim_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\widetilde{\mathbb{P}}^{\varepsilon}_{n,r}\left(\tilde{B}\left(\tilde{A},\delta\right)\right)=-\infty.$$

Let $A\in\mathcal{H}_r^{(\beta)}$ be such that $A\in\mathcal{D}_{\varepsilon,r}$. Since the functions $X\in\mathcal{H}_r^{(\beta)}\mapsto\sum_{i=1}^{r}|X_{i,i}|^{\alpha}$ and $X\in\mathcal{H}_r^{(\beta)}\mapsto\sum_{1\leq i\neq j\leq r}|X_{i,j}|^{\alpha}$ are continuous, then by definition of the topology we equipped $\widetilde{\mathcal{E}}_r$, we deduce that $\tilde{X}\in\widetilde{\mathcal{E}}_r\mapsto\sum_{i\geq 1}|X_{i,i}|^{\alpha}$ and $\tilde{X}\in\widetilde{\mathcal{E}}_r\mapsto\sum_{i\neq j}|X_{i,j}|^{\alpha}$ are continuous. Then, we can find a nonnegative function h, such that $h(\delta)\to 0$ as $\delta\to 0$, and such that

$$\begin{aligned}&\tilde{P}^{\varepsilon}_{n,r}(\tilde{B}(\tilde{A},\delta))\\&\quad\leq\mathbb{P}\Big(\sum_{i\geq 1}|C^{\varepsilon}_{i,i}|^{\alpha}\geq\sum_{i\geq 1}|A_{i,i}|^{\alpha}-h\left(\delta\right),\sum_{i\neq j}|C^{\varepsilon}_{i,j}|^{\alpha}\geq\sum_{i\neq j}|A_{i,j}|^{\alpha}-h\left(\delta\right)\mid C^{\varepsilon}\in\mathcal{E}_r\Big).\end{aligned}$$

But the sets

$$\Big\{\sum_{i\geq 1}|C^{\varepsilon}_{i,i}|^{\alpha}\geq\sum_{i\geq 1}|A_{i,i}|^{\alpha}-h\left(\delta\right)\Big\}\text{ and }\Big\{\sum_{i\neq j}|C^{\varepsilon}_{i,j}|^{\alpha}\geq\sum_{i\geq 1}|A_{i,j}|^{\alpha}-h\left(\delta\right)\Big\},$$

are nondecreasing with respect to the absolute value of each entry of C^{ε}, and $\{C^{\varepsilon}\in\mathcal{E}_r\}$ is nonincreasing with respect to the absolute value of each entry of C^{ε}. Using again Harris' inequality and the independence of the entries,

$$\begin{aligned}\tilde{P}^{\varepsilon}_{n,r}(\tilde{B}(\tilde{A},\delta))&\leq\mathbb{P}\Big(\sum_{i\geq 1}|C^{\varepsilon}_{i,i}|^{\alpha}\geq\sum_{i\geq 1}|A_{i,i}|^{\alpha}-h\left(\delta\right),\sum_{i\neq j}|C^{\varepsilon}_{i,j}|^{\alpha}\geq\sum_{i\neq j}|A_{i,j}|^{\alpha}-h\left(\delta\right)\Big)\\&=\mathbb{P}\Big(\sum_{i\geq 1}|C^{\varepsilon}_{i,i}|^{\alpha}\geq\sum_{i\geq 1}|A_{i,i}|^{\alpha}-h\left(\delta\right)\Big)\mathbb{P}\Big(\sum_{i\neq j}|C^{\varepsilon}_{i,j}|^{\alpha}\geq\sum_{i\neq j}|A_{i,j}|^{\alpha}-h\left(\delta\right)\Big).\end{aligned} \tag{3.59}$$

Let $n \geq r$. By Chernoff's inequality we get, with $0 < b_1 < b$,

$$\mathbb{P}\Big(\sum_{i=1}^{n}|C_{i,i}|^{\alpha} \geq \sum_{i=1}^{n}|A_{i,i}|^{\alpha} + h(\delta)\Big) \leq e^{-n^{\alpha/2}b_1\left(\sum_{i=1}^{n}|A_{i,i}|^{\alpha}+h(\delta)\right)} \times \mathbb{E}\Big(e^{b_1|X_{1,1}|^{\alpha}\mathbb{1}_{\varepsilon n^{1/2}\leq |X_{1,1}|\leq \varepsilon^{-1}n^{1/2}}}\Big)^{n}.$$

But we know from (3.52) that for any $b_2 \in (b_1, b)$ and n large enough,

$$\mathbb{E}\Big(e^{b_1|X_{1,1}|^{\alpha}\mathbb{1}_{\varepsilon n^{1/2}\leq |X_{1,1}|\leq \varepsilon^{-1}n^{1/2}}}\Big) \leq \exp\Big(\frac{b_2}{b_2-b_1}e^{-(b_2-b_1)\varepsilon^{\alpha}n^{\alpha/2}}\Big).$$

Hence,

$$\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\Big(\sum_{i=1}^{n}|C_{i,i}|^{\alpha} \geq \sum_{i=1}^{n}|A_{i,i}|^{\alpha} + h(\delta))\Big) \leq -b_1\sum_{i\geq 1}|A_{i,i}|^{\alpha} + h(\delta).$$

As this inequality is true for all $b_1 < b$, letting b_1 go to b, we get,

$$\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\Big(\sum_{i=1}^{n}|C_{i,i}|^{\alpha} \geq \sum_{i=1}^{n}|A_{i,i}|^{\alpha} + h(\delta)\Big) \leq -b\Big(\sum_{i\geq 1}|A_{i,i}|^{\alpha} + h(\delta)\Big).$$

Similarly one can show,

$$\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\Big(\sum_{i\neq j}|C_{i,j}|^{\alpha} \geq \sum_{i\neq j}|A_{i,j}|^{\alpha} + h(\delta)\Big) \leq -\frac{a}{2}\Big(\sum_{i\neq j}|A_{i,j}|^{\alpha} + h(\delta)\Big).$$

Putting these two last estimates into (3.59), we get

$$\limsup_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\tilde{P}^{\varepsilon}_{n,r}\left(\tilde{B}\left(\tilde{A},\delta\right)\right) \leq -b\sum_{i\geq 1}|A_{i,i}|^{\alpha} - \frac{a}{2}\sum_{i\neq j}|A_{i,j}|^{\alpha}.$$

□

The idea now, is to use the fact that C^{ε} has with exponentially large probability at most r non-zero entries, by Proposition 3.6.7, to release the conditioning on the event $\{C^{\varepsilon} \in \mathcal{E}_r\}$. Then, as the largest eigenvalue map is continuous on $\widetilde{\mathcal{E}}_r$, the contraction principle will give us a LDP for $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$.

3.10.3 Proposition. *Recall that for any $n \in \mathbb{N}$ and $\varepsilon > 0$, we define*

$$\rho_{n,\varepsilon} = \begin{cases} g_{\mu_{sc}}^{-1}\left(1/\lambda_{C^{\varepsilon}}\right) & \text{if } \lambda_{C^{\varepsilon}} \geq 1, \\ 2 & \text{otherwise,}\end{cases}$$

where $\lambda_{C^{\varepsilon}}$ denotes the largest eigenvalue of C^{ε}, and C^{ε} is as in (3.6).

For all $\varepsilon > 0$, $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ follows a LDP with speed $n^{\alpha/2}$, and good rate function J_{ε}, defined by

$$J_{\varepsilon}(x) = \begin{cases} \inf\{I_{\varepsilon}(A) : A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)},\ \lambda_A = 1/g_{\mu_{sc}}(x)\} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2,\end{cases}$$

where λ_A denotes the largest eigenvalue of any Hermitian matrix A and

$$I_\varepsilon(A) = \begin{cases} b\sum_{i\geq 1}|A_{i,i}|^\alpha + a\sum_{i<j}|A_{i,j}|^\alpha & \text{if } A \in \mathcal{D}_\varepsilon, \\ +\infty & \text{otherwise.} \end{cases}$$

with

$$\mathcal{D}_\varepsilon = \left\{A \in \cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)} : \forall i \leq j,\ A_{i,j} = 0 \text{ or } \varepsilon \leq |A_{i,j}| \leq \varepsilon^{-1}, A_{i,j}/|A_{i,j}| \in \operatorname{supp}(\nu_{i,j})\right\},$$

with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in 3.2.1.

Proof. Note that by Lemma 3.6.5, we already know that $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ is exponentially tight. Therefore, we only need to show that $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ satisfies a weak LDP. Let $f : \cup_{m\geq 1}\mathcal{H}_m^{(\beta)} \to \mathbb{R}$ be defined by,

$$f(A) = \begin{cases} g_{\mu_{sc}}^{-1}(1/\lambda_A) & \text{if } \lambda_A \geq 1, \\ 2 & \text{otherwise.} \end{cases}$$

Since the largest eigenvalue of a Hermitian matrix is invariant by conjugation, f can be defined on $\widetilde{\mathcal{E}}_r$ for any $r \in \mathbb{N}$. Because of the topology we put on $\widetilde{\mathcal{E}}_r$, f is continuous on $\widetilde{\mathcal{E}}_r$. Therefore, by the contraction principle (see [43, Theorem 4.2.1]), the push-forward of $\widetilde{\mathbb{P}}_{n,r}^{\varepsilon}$ by f, denoted $\widetilde{\mathbb{P}}_{n,r}^{\varepsilon} \circ f^{-1}$, satisfies a LDP with speed $n^{\alpha/2}$, and good rate function $J_{\varepsilon,r}$, defined for any $x \in \mathbb{R}$ by

$$J_{\varepsilon,r}(x) = \inf\left\{I_{\varepsilon,r}(\tilde{A}) : f(\tilde{A}) = x, \tilde{A} \in \widetilde{\mathcal{E}}_r\right\},$$

where $I_{\varepsilon,r}$ is as in (3.49). Since $g_{\mu_{sc}}^{-1}(x) \geq 2$, for all $x \in (0,1]$, we can re-write this rate function as,

$$J_{\varepsilon,r}(x) = \begin{cases} \inf\{I_\varepsilon(A) : \lambda_A = 1/g_{\mu_{sc}}(x), A \in \mathcal{D}_\varepsilon\} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2, \end{cases}$$

where I_ε and $\mathcal{D}_\varepsilon$ are defined in Proposition 3.10.3. Observe that $\widetilde{\mathbb{P}}_{n,r}^{\varepsilon} \circ f^{-1}$ is in fact the law of $\rho_{n,\varepsilon}$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$. We will show that $(\widetilde{\mathbb{P}}_{n,r}^{\varepsilon} \circ f^{-1})_{n,r\in\mathbb{N}}$ are exponentially good approximations of $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$. Let $\nu_{r,n}$ be an independent random variable with the same law as of $\rho_{n,\varepsilon}$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$. Define

$$\tilde{\nu}_{r,n} = \rho_{n,\varepsilon}\mathbb{1}_{C^\varepsilon\in\mathcal{E}_r} + \nu_{r,n}\mathbb{1}_{C^\varepsilon\notin\mathcal{E}_r}.$$

Then, $\tilde{\nu}_{r,n}$ and $\nu_{r,n}$ have the same law $\widetilde{\mathbb{P}}_{n,r}^{\varepsilon} \circ f^{-1}$. Let $\delta > 0$. We have

$$\mathbb{P}\left(|\tilde{\nu}_{r,n} - \rho_{n,\varepsilon}| > \delta\right) \leq \mathbb{P}\left(C^\varepsilon \notin \mathcal{E}_r\right).$$

By Proposition 3.6.7, we get

$$\lim_{r\to+\infty}\limsup_{n\to+\infty}\frac{1}{n^{\alpha/2}}\log\mathbb{P}\left(|\tilde{\nu}_{r,n} - \rho_{n,\varepsilon}| > \delta\right) = -\infty.$$

We can apply [43, Theorem 4.2.16] and deduce that $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ satisfies a weak LDP with speed $n^{\alpha/2}$, and rate function defined for all $x \in \mathbb{R}$ by

$$\Psi_\varepsilon(x) = \sup_{\delta>0} \liminf_{r\to+\infty} \inf_{|x-y|<\delta} J_{\varepsilon,r}(y).$$

But $J_{\varepsilon,r}$ is nonincreasing in r. Thus,

$$\Psi_\varepsilon(x) = \sup_{\delta>0} \inf_{r>0} \inf_{|x-y|<\delta} J_{\varepsilon,r}(y) = \sup_{\delta>0} \inf_{|x-y|<\delta} \inf_{r>0} J_{\varepsilon,r}(y) = \sup_{\delta>0} \inf_{|x-y|<\delta} J_\varepsilon(y),$$

where J_ε is defined in Proposition 3.10.3. To conclude that $\Psi_\varepsilon = J_\varepsilon$, we need to show that J_ε is lower semicontinuous. We will in fact show that J_ε has compact level sets. Let $\tau > 0$ and $x \in \mathbb{R}$. If

$$\inf\{I_\varepsilon(A) : f(A) = x, A \in \cup_{m\in\mathbb{N}} \mathcal{H}_m^{(\beta)}\} \leq \tau,$$

where I_ε is defined in Proposition 3.10.3, then

$$\inf\{I_\varepsilon(A) : f(A) = x\} = \inf\{I_\varepsilon(A) : f(A) = x,\ I_\varepsilon(A) \leq 2\tau\}$$

But if $A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)}$ is such that $I_\varepsilon(A) \leq 2\tau$, then

$$\big(b \wedge \frac{a}{2}\big) \sum_{i,j} \varepsilon^\alpha \mathbb{1}_{A_{i,j}\neq 0} \leq I_\varepsilon(A) \leq 2\tau.$$

Let $r \geq \frac{2\tau}{\varepsilon^\alpha (b\wedge a/2)}$. We deduce by the above observation that,

$$\inf\{I_\varepsilon(A) : f(A) = x\} = \inf\{I_\varepsilon(A) : f(A) = x, I_\varepsilon(A) \leq 2\tau, A \in \mathcal{E}_r\}.$$

Therefore,

$$\inf\{I_\varepsilon(A) : f(A) = x\} = \inf\{I_\varepsilon(A) : f(A) = x, A \in \mathcal{E}_r\}.$$

Thus,

$$\inf\{I_\varepsilon(A) : f(A) = x\} = \inf\left\{I_{\varepsilon,r}\left(\tilde{A}\right) : f(\tilde{A}) = x, \tilde{A} \in \tilde{\mathcal{E}}_r\right\},$$

with $I_{\varepsilon,r}$ being defined in Proposition 3.10.1. Since $I_{\varepsilon,r}$ is a good rate function and f is continuous on $\tilde{\mathcal{E}}_r$, we have

$$\{x \in \mathbb{R} : J_\varepsilon(x) \leq \tau\} = \left\{f(\tilde{A}) : I_{\varepsilon,r}\left(\tilde{A}\right) \leq \tau\right\}.$$

Thus, the τ-level set of J_ε is compact, which concludes the proof.

□

We are now ready to give a proof the main result of this chapter. More precisely, we will prove the following.

3.10.4 Theorem. *The sequence $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ follows a LDP with speed $n^{\alpha/2}$, and good rate function defined by,*

$$J(x) = \begin{cases} c g_{\mu_{sc}}(x)^{-\alpha} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2, \end{cases}$$

where

$$c = \inf \left\{ W_\alpha(A) : \lambda_A = 1, A \in \mathcal{D} \right\}, \tag{3.60}$$

where W_α *is defined for any* $A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)}$, *by*

$$W_\alpha(A) = b\sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + a\sum_{i<j} |A_{i,j}|^\alpha,$$

and

$$\mathcal{D} = \left\{ A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)} : \forall i \leq j,\ A_{i,j} = 0 \text{ or } \frac{A_{i,j}}{|A_{i,j}|} \in \operatorname{supp}(\nu_{i,j}) \right\},$$

where $\nu_{i,j} = \nu_1$ *if* $i = j$, *and* ν_2 *if* $i < j$, *and where* $\operatorname{supp}(\nu_{i,j})$ *denotes the support of the measure* $\nu_{i,j}$.

Proof. We already know by Proposition 3.6.1 that $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ is exponentially tight. Thus, it is sufficient to prove that $(\lambda_{X/\sqrt{n}})_{\in\mathbb{N}}$ satisfies a weak LDP. Since we know from Theorem 3.9.1 that $(\mu_{\varepsilon,n})_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$, and that for each $\varepsilon > 0$, $(\rho_{n,\varepsilon})_{n\in\mathbb{N}}$ follows a LDP with rate function J_ε, then by [43, Theorem 4.2.16], we deduce that $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$, satisfies a weak LDP with rate function,

$$\Phi(x) = \sup_{\delta>0} \liminf_{\varepsilon\to 0} \inf_{|y-x|<\delta} J_\varepsilon(y),$$

As J_ε is nondecreasing in ε, we get

$$\begin{aligned}\Phi(x) &= \sup_{\delta>0} \inf_{\varepsilon>0} \inf_{|y-x|<\delta} J_\varepsilon(y) = \sup_{\delta>0} \inf_{|y-x|<\delta} \inf_{\varepsilon>0} J_\varepsilon(y)\\ &= \sup_{\delta>0} \inf_{|y-x|<\delta} J(x), \end{aligned} \tag{3.61}$$

with

$$J(x) = \begin{cases} \inf \left\{ W_\alpha(A) : A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)}, \lambda_A = g_{\mu_{sc}}(x)^{-1}, A \in \mathcal{D} \right\} & \text{if } x > 2, \\ 0 & \text{if } x = 2, \\ +\infty & \text{if } x < 2. \end{cases} \tag{3.62}$$

As for any $t > 0$, and $A \in \mathcal{H}_m^{(\beta)}$, $W_\alpha(tA) = t^\alpha W_\alpha(A)$ and $\lambda_{tA} = t\lambda_A$, and furthermore $\mathcal{D}$ is a cone, we have for any $x > 2$,

$$J(x) = g_{\mu_{sc}}(x)^{-\alpha} J(1).$$

As $g_{\mu_{sc}}$ is non-increasing from $[2, +\infty)$ to $(0, 1]$. This yields that J has compact level sets. Therefore, from (3.61), we get that $\Phi = J$, which concludes the proof. □

3.11 Appendix: computation of $J(1)$

In this section, we compute the constant c in Theorem 3.10.4 explicitly, assuming certain conditions on the supports of the limiting angle distributions of the diagonal and off-diagonal entries (in the sense of 3.2.1). In particular, when the entries are real random variables, or when $\alpha \in (0, 1]$, the following proposition together with Theorem 3.10.4, gives an explicit formula for the rate function.

3.11.1 Proposition. *With the notations of Theorem 3.10.4, we have the following:*
(a). If $0 < \alpha \leq 1$, then

$$c = \begin{cases} \min(b, a) & \text{if } 1 \in \operatorname{supp}(\nu_1), \\ a & \text{otherwise.} \end{cases}$$

(b). If $1 < \alpha < 2$ and $1 \in \operatorname{supp}(\nu_1)$, and $b \leq \frac{a}{2}$, then $c = b$.
(c). If $1 < \alpha < 2$, $1 \in \operatorname{supp}(\nu_1) \cap \operatorname{supp}(\nu_2)$ and $b > \frac{a}{2}$, then

$$c = \min\Big\{W_\alpha\big(P_k(p, q)\big) : k \in \mathbb{N}\Big\},$$

with $p = b^{-\frac{1}{\alpha-1}}$ and $q = (a/2)^{-\frac{1}{\alpha-1}}$, where $P_k(s,t)$ denotes for any $(s,t) \neq (0,0)$, and $k \in \mathbb{N}$, the following matrix of size $k \times k$,

$$P_k(s,t) = \frac{1}{s + (k-1)t} \begin{pmatrix} s & t & \cdots & t \\ t & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & t \\ t & \cdots & t & s \end{pmatrix}. \tag{3.63}$$

Equivalently,

$$c = \min\left(\psi\left(\lfloor t_0 \rfloor\right), \psi\left(\lceil t_0 \rceil\right)\right),$$

where $\lfloor t_0 \rfloor$ and $\lceil t_0 \rceil$ denote respectively the lower and upper integer parts of t_0, and with ψ and t_0 being defined by

$$\forall t \geq 1, \ \psi(t) = \frac{t}{\left(p + (t-1)\, q\right)^{(\alpha-1)}}, \ t_0 = \frac{1}{2-\alpha}\Big(1 - \frac{p}{q}\Big). \tag{3.64}$$

(d). If $1 < \alpha < 2$, $1 \in \operatorname{supp}(\nu_1)$, and $\operatorname{supp}(\nu_2) = \{-1\}$ and $b > \frac{a}{2}$, then,

$$c = \min\Big(b, \frac{2}{(p+q)^{\alpha-1}}\Big).$$

(e). If $1 < \alpha < 2$, $\operatorname{supp}(\nu_1) = \{-1\}$ and $1 \in \operatorname{supp}(\nu_2)$, then

$$c = \min\left\{W_\alpha(P_k\left(0,1\right)) : k \geq 2\right\} = \frac{a}{2} \min\left(\varphi\left(\lfloor t_1 \rfloor\right), \varphi\left(\lceil t_1 \rceil\right)\right),$$

where

$$\forall t \geq 2, \ \varphi(t) = \frac{t}{(t-1)^{\alpha-1}}, \quad t_1 = \frac{1}{2-\alpha}.$$

(f). If $1 < \alpha < 2$, and $\operatorname{supp}(\nu_1) = \operatorname{supp}(\nu_2) = \{-1\}$, then $c = a$.

Proof. (a). Let $0 < \alpha \leq 1$ and $1 \in \operatorname{supp}(\nu_1)$. Note that if $A \in \mathcal{H}_n^{(\beta)}$ is such that $|A_{i,j}| \geq 1$, for some $i, j \in \{1, ..., n\}$, then $W_\alpha(A) \geq \min(b, a)$. But, as $1 \in \operatorname{supp}(\nu_1)$,

$$c \leq \min\left(W_\alpha\left(1\right), W_\alpha\begin{pmatrix} 0 & e^{i\theta} \\ e^{-i\theta} & 0 \end{pmatrix}\right), \tag{3.65}$$

with some $\theta \in \mathrm{supp}(\nu_2)$. Therefore $c \leq \min(b, a)$. We deduce that we can restrict the constraints set of the optimization problem (3.60) to matrices with entries less or equal than 1 in absolute value. As $0 < \alpha \leq 1$, we get,

$$\begin{aligned} c &\geq (b \wedge a) \inf \Big\{ \sum_{i \geq 1} |A_{i,i}| + \sum_{i<j} |A_{i,j}| : \lambda_A = 1, A \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)} \Big\} \\ &\geq (b \wedge a) \inf \Big\{ \frac{1}{2} |\mathrm{tr}(A)| + \frac{1}{2} \sum_{i,j} |A_{i,j}| : \lambda_A = 1, A \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)} \Big\}, \end{aligned}$$

where used the triangular inequality in the last inequality. But we know from [104, Theorem 3.32], that for any $A \in \mathcal{H}_n^{(\beta)}$,

$$\sum_{i,j} |A_{i,j}| \geq \sum_{i=1}^{n} |\lambda_i| \,,$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A. Therefore,

$$c \geq \frac{1}{2} (b \wedge a) \inf_{n \geq 1} \inf \Big\{ \big|1 + \sum_{i=1}^{n-1} \lambda_i \big| + \big(1 + \sum_{i=1}^{n-1} |\lambda_i| \,\big) : \lambda_1, ..., \lambda_{n-1} \in \mathbb{R} \Big\}.$$

But, for all $\lambda_1, ..., \lambda_{n-1} \in \mathbb{R}$,

$$\big|1 + \sum_{i=1}^{n-1} \lambda_i \big| + \big(1 + \sum_{i=1}^{n-1} |\lambda_i| \,\big) \geq 2 + \sum_{i=1}^{n-1} (\lambda_i + |\lambda_i|) \geq 2,$$

with equality for $\lambda_1 = \lambda_2 = ... = \lambda_{n-1} = 0$. We conclude that $c = \min(b, a)$.

Let $0 < \alpha \leq 1$, but assume $\mathrm{supp}(\nu_1) = \{-1\}$. Then,

$$\begin{aligned} c &\geq \inf \Big\{ b \sum_{i \geq 1} |A_{i,i}| + a \sum_{i<j} |A_{i,j}| : A \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)}, A_{i,i} \leq 0, \forall i \in \mathbb{N}, \lambda_A = 1 \Big\} \\ &= \inf \Big\{ \Big(b - \frac{a}{2}\Big) \big| \sum_{i \geq 1} A_{i,i} \big| + \frac{a}{2} \sum_{i,j} |A_{i,j}| : A \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)}, A_{i,i} \leq 0, \forall i \in \mathbb{N}, \lambda_A = 1 \Big\} \\ &\geq \inf \Big\{ \Big(b - \frac{a}{2}\Big) \big| \sum_{i \geq 1} A_{i,i} \big| + \frac{a}{2} \sum_{i,j} |A_{i,j}| : A \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)}, \mathrm{tr} A \leq 0, \forall i \in \mathbb{N}, \lambda_A = 1 \Big\}. \end{aligned}$$

Using again the fact that $\sum_{i,j} |A_{i,j}| \geq \sum_{i=1}^{n} |\lambda_i|$, where $A \in \mathcal{H}_n^{(\beta)}$, and $\lambda_1, ..., \lambda_n$, are the eigenvalues of A, we get

$$c \geq \inf_{n \geq 1} \inf \Big\{ \Big(b - \frac{a}{2}\Big) \big|1 + \sum_{i=1}^{n-1} \lambda_i \big| + \frac{a}{2} \big(1 + \sum_{i=1}^{n-1} |\lambda_i| \,\big) : \lambda_1, ..., \lambda_{n-1} \in \mathbb{R}, \sum_{i=1}^{n-1} \lambda_i \leq -1 \Big\}.$$

But if $1 + \sum_{i=1}^{n-1} \lambda_i \leq 0$, for $\lambda_1, ..., \lambda_{n-1} \in \mathbb{R}$, then

$$\begin{aligned} \Big(b - \frac{a}{2}\Big) \Big|1 + \sum_{i=1}^{n-1} \lambda_i \Big| + \frac{a}{2} \big(1 + \sum_{i=1}^{n-1} |\lambda_i| \,\big) &= -\Big(b - \frac{a}{2}\Big) \big(1 + \sum_{i=1}^{n-1} \lambda_i\big) + \frac{a}{2} \big(1 + \sum_{i=1}^{n-1} |\lambda_i| \,\big) \\ &= a - b\big(1 + \sum_{i=1}^{n-1} \lambda_i\big) + \frac{a}{2} \sum_{i=1}^{n-1} (|\lambda_i| + \lambda_i) \\ &\geq a. \end{aligned}$$

Thus, $c \geq a$. But, $c \leq a$ by the same argument as in (3.65), therefore $c = a$.

(b). Let $1 < \alpha < 2$ and assume $1 \in \mathrm{supp}(\nu_1)$ and $b \leq \frac{a}{2}$. Due to [104, Theorem 3.32], we have for any $A \in \mathcal{H}_n^{(\beta)}$,

$$W_\alpha(A) \geq b \sum_{1 \leq i,j \leq n} |A_{i,j}|^\alpha \geq b \sum_{i=1}^n |\lambda_i|^\alpha ,$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A. As $\lambda_A = 1$, we get $W_\alpha(A) \geq b$. Therefore, $c \geq b$. As $1 \in \mathrm{supp}(\nu_1)$, we also have $c \leq W_\alpha((1)) = b$, which ends the proof.

(c). Let $1 < \alpha < 2$, $b > \frac{a}{2}$ and assume $1 \in \mathrm{supp}(\nu_1) \cap \mathrm{supp}(\nu_2)$. We have the bound

$$c \geq \inf_{n \geq 1} \inf \left\{ W_\alpha(A) : A \in \mathcal{H}_n^{(\beta)}, \lambda_A = 1 \right\}.$$

Let $m \geq 2$. We consider the minimization problem

$$\inf \left\{ W_\alpha(A) : A \in \mathcal{H}_n^{(\beta)}, \lambda_A = 1 \right\}.$$

As I is continuous and the constraints set is compact, the infimum is achieved at some $A \in \mathcal{H}_n^{(\beta)}$. If 1 is an eigenvalue of A of multiplicity greater that 2, then denoting by $\lambda_1, ..., \lambda_n$ the eigenvalues of A, we have by [104, Theorem 3.32],

$$W_\alpha(A) \geq \frac{a}{2} \sum_{i=1}^n |\lambda_i|^\alpha \geq a.$$

As A is a minimizer, and $1 \in \mathrm{supp}(\nu_1) \cap \mathrm{supp}(\nu_2)$,

$$W_\alpha(A) \leq W_\alpha \begin{pmatrix} p & (1-p) \\ (1-p) & p \end{pmatrix},$$

where $p = \left(1 + \left(\frac{2b}{a}\right)^{1/(\alpha-1)}\right)^{-1}$. As $2bp^{\alpha-1} = a(1-p)^{\alpha-1}$,

$$\begin{aligned} W_\alpha \begin{pmatrix} p & (1-p) \\ (1-p) & p \end{pmatrix} &= 2bp^\alpha + a(1-p)^\alpha \\ &= a(1-p)^{\alpha-1}p + a(1-p)^\alpha \\ &= a(1-p)^{\alpha-1} < a, \end{aligned}$$

where we used in the last inequality the fact that $\alpha > 1$. This yields a contradiction.

Therefore, 1 must be a simple eigenvalue of A. From the multipliers rule (see [40, Theorem 10.48]), there exist $\eta, \gamma \in \mathbb{R}$, $(\eta, \gamma) \neq 0$, such that $\eta = 0$, or 1, and

$$0 \in \eta\{\nabla W_\alpha(A)\} - \gamma \partial \lambda(A), \tag{3.66}$$

where the gradient of f, and the subdifferential of λ, denoted $\partial\lambda$, are taken with respect to the canonical Hermitian product on $\mathcal{H}_n^{(\beta)}$. As a consequence of Danskin's formula (see [40, Theorem 10.22]), we have the following lemma.

3.11.2 Lemma. *Let $\lambda : \mathcal{H}_n^{(\beta)} \to \mathbb{R}$ be the largest eigenvalue function. The subdifferential of λ at A, taken with respect to the canonical Hermitian product, is*

$$\partial\lambda(A) = \left\{ X \in \mathcal{H}_n^{(\beta)} : 0 \leq X \leq \mathbb{1}_{E_{\lambda_A}(A)}, \mathrm{tr}X = 1 \right\},$$

where $\mathbb{1}_{E_{\lambda_A}(A)}$ denotes the projection on the eigenspace $E_{\lambda_A}(A)$ of A associated with the largest eigenvalue of A, and $\leq$ is the order structure on $\mathcal{H}_n^{(\beta)}$.

As 1 is a simple eigenvalue of A, we get from Lemma 3.11.2 that there is some unit eigenvector of A, x, associated with the eigenvalue 1, such that

$$\eta \nabla W_\alpha(A) = \gamma x x^*.$$

We deduce that for any $i \neq j$,

$$\eta \frac{a}{2} \alpha A_{i,j} \left| A_{i,j} \right|^{\alpha-2} = \gamma x_i \overline{x_j}, \tag{3.67}$$

and for any $1 \leq i \leq n$,

$$\eta b \alpha A_{i,i} \left| A_{i,i} \right|^{\alpha-2} = \gamma \left| x_i \right|^2, \tag{3.68}$$

with the convention that $z|z|^{\alpha-2} = 0$ when $z = 0$. Multiplying the two equations above by $\overline{A_{i,j}}$ and $A_{i,i}$ respectively, and summing over all $i, j \in \{1, ..., n\}$, we get

$$\eta W_\alpha(A) = \gamma. \tag{3.69}$$

As $(\eta, \gamma) \neq (0, 0)$, this shows that $\eta = 1$. Furthermore, the stationary condition yields for all $i \neq j$,

$$A_{i,j} = \left(\frac{2\gamma}{a\alpha} \right)^{\frac{1}{\alpha-1}} x_i \overline{x_j} \left| x_i x_j \right|^{\frac{1}{\alpha-1}-1},$$

and for all $1 \leq i \leq n$,

$$A_{i,i} = \left(\frac{\gamma}{b\alpha} \right)^{\frac{1}{\alpha-1}} \left| x_i \right|^{\frac{2}{\alpha-1}}.$$

Due to the eigenvalue equation $Ax = x$, we have for all $1 \leq i \leq n$,

$$\left(\frac{\gamma}{b\alpha} \right)^{\frac{1}{\alpha-1}} \left| x_i \right|^{\frac{2}{\alpha-1}} x_i + \left(\frac{2\gamma}{a\alpha} \right)^{\frac{1}{\alpha-1}} \sum_{j \neq i} x_i \left| x_i \right|^{\frac{1}{\alpha-1}-1} \left| x_j \right|^{\frac{1}{\alpha-1}+1} = x_i. \tag{3.70}$$

At the price of permuting the coordinates of x and conjugating A by a permutation matrix, we can assume $x = (x_1, ..., x_k, 0, ..., 0)$, with $x_1 \neq 0, ..., x_k \neq 0$. Dividing by $x_i |x_i|^{\frac{1}{\alpha-1}-1}$ in (3.70), we get

$$By = \left(\frac{\gamma}{\alpha} \right)^{-\frac{1}{\alpha-1}} y^{-\frac{2-\alpha}{\alpha}}, \tag{3.71}$$

where $y \in \mathbb{R}^k$ is such that $y_i = |x_i|^{1+\frac{1}{\alpha-1}}$ for all $i \in \{1, ..., k\}$, and where the power on the right-hand side must be understood entry-wise, and $B = P_k(p, q)$, with $P_k(p, q)$ defined in (3.63) with $p = b^{-1/(\alpha-1)}$ and $q = (2/a)^{1/(\alpha-1)}$. As x is a unit vector, we have $\sum_{i=1}^k y_i^{2(\alpha-1)/\alpha} = 1$. Taking the scalar product with y in (3.71), yields

$$\left(\frac{\gamma}{\alpha} \right)^{-\frac{1}{\alpha-1}} = \langle By, y \rangle.$$

As $I(A) = \frac{\gamma}{\alpha}$ by (3.69), we deduce that

$$c \geq \left(\sup_{k\geq 1} \sup \left\{ \langle By, y\rangle : y \in [0,+\infty)^k, \sum_{i=1}^k y_i^{2(\alpha-1)/\alpha} \right\} \right)^{-(\alpha-1)}. \tag{3.72}$$

In the next lemma, we compute the maximum of certain quadratic forms, like the one given by the matrix B, on the unit ℓ^δ-sphere, intersected with $[0,+\infty)^n$, where $\delta \in (0,1)$.

3.11.3 Lemma. *Let $\lambda, \mu \in \mathbb{R}$ such that $0 \leq \lambda < \mu$. Let $\delta \in (0,1)$. Define for any $n \in \mathbb{N}$,*

$$B = \begin{pmatrix} \lambda & \mu & \cdots & \mu \\ \mu & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mu \\ \mu & \cdots & \mu & \lambda \end{pmatrix} \in \mathcal{H}_n^{(1)}.$$

It holds

$$\sup \left\{ \langle By, y\rangle : y \in [0,+\infty)^n, \sum_{i=1}^n y_i^\delta = 1 \right\} = \max_{1\leq k\leq n} (\lambda + (k-1)\mu)k^{1-2/\delta}. \tag{3.73}$$

Proof. Let $n \in \mathbb{N}$. By continuity and compactness arguments, we see that the supremum

$$\sup \left\{ \langle By, y\rangle : y \in [0,+\infty)^n, \sum_{i=1}^n y_i^\delta = 1 \right\},$$

is achieved at some $y \in \mathbb{R}^n$. At the price of re-ordering the coordinates of y, we can assume that $y = (z_1, ..., z_m, 0..., 0)$, with $z_1 > 0, ..., z_m > 0$, for some $m \in \{1, ..., n\}$. Then, the vector $z = (z_1, ..., z_m) \in \mathbb{R}^m$ is a solution of the optimization problem

$$\sup \left\{ \langle Bz, z\rangle : z \in [0,+\infty)^m, \sum_{i=1}^m z_i^\delta = 1 \right\},$$

which lies in the interior of $[0,+\infty)^m$. The multipliers rule (see [40][Theorem 9.1]) asserts that there is some $(\eta, \gamma) \neq (0,0)$, with $\eta = 0$ or 1, such that

$$\eta Bz = \gamma z^{\delta-1}, \tag{3.74}$$

where the power on the right-hand side has to be understood entry-wise. Taking the scalar product with z in (3.74) yields

$$\eta \langle Bz, z\rangle = \gamma.$$

We deduce that $\eta = 1$. Moreover, by (3.74) we have for any $i \in \{1, ..., m\}$,

$$\mu \sum_{j=1}^m z_j = \gamma z_i^{\delta-1} + (\mu - \lambda) z_i. \tag{3.75}$$

But then, the function

$$\forall s \in (0,+\infty),\ f(s) = \gamma s^{\delta-1} + (\mu-\lambda)s,$$

is decreasing on $(0, s_0]$, and increasing on $(s_0, +\infty)$, for some $s_0 \in (0,+\infty)$. Thus, (3.75) yields that z has at most two distinct coordinates. Thus, there are some $k, l \in \mathbb{N}$, such that $k + l = m$, and $s, t \geq 0$, such that $ks^\delta + lt^\delta = 1$, and

$$\forall i \in \{1, ..., m\},\ z_i = \mathbb{1}_{i\leq k} s + \mathbb{1}_{k+1\leq i\leq k+l} t.$$

But then,

$$\begin{aligned}\langle Bz, z\rangle &= \lambda k s^2 + \mu k(k-1)s^2 + \lambda l t^2 + \mu l(l-1)t^2 + 2\mu k l s t\\ &= k\left(\lambda + (k-1)\mu\right)s^2 + l\left(\lambda + (l-1)\mu\right)t^2 + 2\mu k l s t.\end{aligned}$$

Let $x = ks^\delta$. We get

$$\begin{aligned}\langle Bz, z\rangle = k^{1-2/\delta}\left(\lambda + (k-1)\mu\right)x^{2/\delta} + l^{1-2/\delta}\left(\lambda + (l-1)\mu\right)(1-x)^{2/\delta}&\\ +2\mu(kl)^{1-1/\delta}x^{1/\delta}(1-x)^{1/\delta}&.\end{aligned}$$

Define

$$\forall x \in (0,+\infty),\ \varphi(x) = x^{1-2/\delta}(\lambda + (x-1)\mu).$$

Note that φ is increasing on $(0, x_0]$ and decreasing on $[x_0, +\infty)$, where

$$x_0 = \frac{(\frac{2}{\delta}-1)}{\frac{2}{\delta}-2}\Big(1 - \frac{\lambda}{\mu}\Big). \tag{3.76}$$

With this definition, we have

$$\langle Bz, z\rangle = \varphi(k)x^{2/\delta} + \varphi(l)(1-x)^{2/\delta} + 2\mu(kl)^{1-1/\delta}x^{1/\delta}(1-x)^{1/\delta}.$$

Therefore,

$$\max\Big\{\langle Bz, z\rangle : z \in [0,+\infty)^m, \sum_{i=1}^m z_i^\delta = 1\Big\} = \max_{\substack{k+l\leq m\\ k,l\in\mathbb{N}}}\max_{x\in[0,1]} f_{k,l}(x),$$

with

$$\forall x \in [0,1],\ f_{k,l}(x) = \varphi(k)x^{2/\delta} + \varphi(l)(1-x)^{2/\delta} + 2\mu(kl)^{1-1/\delta}x^{1/\delta}(1-x)^{1/\delta}.$$

Let $m \in \mathbb{N}$, be such that $\varphi(m) = \max\{\varphi(k) : k \in \mathbb{N}^*\}$. Since φ is increasing on $(0, x_0]$ and decreasing on $[x_0, +\infty)$, we have $m \in \{\lfloor x_0\rfloor, \lceil x_0\rceil\}$. Moreover φ, restricted on $\mathbb{N}\backslash\{0\}$, is increasing on $\{1, ..., m\}$, and decreasing on $\{m, m+1, ..., n\}$. As $\delta \in (0,1)$, we have for any $k, l \in \mathbb{N}$, and $x \in [0,1]$,

$$f_{k,l}(x) \leq \varphi(k\wedge m)x^{2/\delta} + \varphi(l\wedge m)(1-x)^{2/\delta} + 2\mu((k\wedge m)(l\wedge m))^{1-1/\delta}x^{1/\delta}(1-x)^{1/\delta}.$$

Therefore,

$$\max\Big\{\langle By, y\rangle : y \in [0,+\infty)^m \sum_{i=1}^m y_i^\delta = 1\Big\} = \max_{\substack{k+l\leq m\\ k,l\leq m}}\max_{x\in[0,1]} f_{k,l}(x). \tag{3.77}$$

We are reduced to study the maximum of certain functions $f_{k,l}$ on the interval $[0,1]$. The variations of those functions are given by the following lemma.

3.11.4 Lemma. *Let $a, b, c \geq 0$, $a, c \neq 0$. Let also $\delta \in (0, 1)$. Define*

$$\forall x \in [0,1], \;\; f(x) = ax^{2/\delta} + 2bx^{1/\delta}(1-x)^{1/\delta} + c(1-x)^{2/\delta}.$$

Then one of the following holds :
(a). There is some $x_1 \in (0,1)$, such that f is decreasing on $[0, x_1]$, and increasing on $[x_1, 1]$.
(b). There are some $0 < x_1 < x_2 < x_3 < 1$, such that f is decreasing on $[0, x_1]$ and $[x_2, x_3]$, and increasing on $[x_1, x_2]$ and $[x_3, 1]$.

Proof. We have, for all $x \in (0, 1)$,

$$\frac{\delta}{2}f'(x) = ax^{\frac{2}{\delta}-1} + bx^{\frac{1}{\delta}-1}(1-x)^{\frac{1}{\delta}} - bx^{\frac{1}{\delta}}(1-x)^{\frac{1}{\delta}-1} - c(1-x)^{\frac{2}{\delta}-1}.$$

We write

$$\frac{\delta}{2}f'(x) = x^{\frac{2}{\delta}-1}\left(a + bs^{\frac{1}{\delta}} - bs^{\frac{1}{\delta}-1} - cs^{\frac{2}{\delta}-1}\right), \tag{3.78}$$

where $s = \frac{1-x}{x}$. Set for all $s \in (0, +\infty)$, $g(s) = a + bs^{\frac{1}{\delta}} - bs^{\frac{1}{\delta}-1} - cs^{\frac{2}{\delta}-1}$. Then, for any $s \in (0, +\infty)$

$$g'(s) = \frac{b}{\delta}s^{\frac{1}{\delta}-1} - b\Big(\frac{1}{\delta} - 1\Big)s^{\frac{1}{\delta}-2} - c\Big(\frac{2}{\delta} - 1\Big)s^{\frac{2}{\delta}-2} = s^{\frac{1}{\delta}-2}h(s),$$

with $h(s) = \frac{b}{\delta}s - b(\frac{1}{\delta} - 1) - c(\frac{2}{\delta} - 1)s^{\frac{1}{\delta}}$. Deriving once more, we get for any $s \in (0, +\infty)$,

$$h'(s) = \frac{b}{\delta} - \frac{c}{\delta}\Big(\frac{2}{\delta} - 1\Big)s^{\frac{1}{\delta}-1}.$$

As $\delta \in (0, 1)$, we see that h' is decreasing. This entails that f has at most three changes of variations. As $f'(0) < 0$, and $f'(1) < 0$, we deduce that f is either decreasing on $[0, x_1]$, and increasing on $[x_1, 1]$, for some $x_1 \in [0, 1]$, or there are some $x_1 < x_2 < x_3$ such that f is decreasing on $[0, x_1]$ and $[x_2, x_3]$, and increasing on $[x_1, x_2]$ and $[x_3, 1]$.

□

We come back at the proof of Lemma 3.11.3. Let $k, l \in \mathbb{N}$, such that $k + l \leq m$ and $1 \leq k \leq l \leq m$. If $k = l$, then $f_{k,l}(x) = f_{k,l}(1 - x)$ for any $x \in [0, 1]$. By Lemma 3.11.4, this entails that if $f_{k,l}$ has a local maximum in $(0, 1)$, then it must be at $1/2$. One can easily check that $f_{k,k}(1/2) = \varphi(2k)$. Thus,

$$\max_{x\in[0,1]} f_{k,k}(x) = \max\left(\varphi(2k), \varphi(k)\right).$$

This shows also that for $m = 1$, we can compute the right-hand side of (3.77).

Assume now $m \geq 2$, and $1 \leq k < l \leq m$. We will show that the maximum of $f_{k,l}$ is achieved at either 0, $k/(k + l)$ or 1. We can write for any $x \in [0, 1]$,

$$\frac{\delta}{2}f'_{k,l}(x) = \left(\frac{x(1-x)}{kl}\right)^{\frac{1}{\delta}-\frac{1}{2}} g_{k,l}(y), \tag{3.79}$$

with $y = \frac{k(1-x)}{lx}$, and $g_{k,l}(y) = (\lambda + (k-1)\mu)y^{-\frac{1}{\delta}+\frac{1}{2}} + \mu\left(ly^{\frac{1}{2}} - ky^{-\frac{1}{2}}\right) - (\lambda + (l-1)\mu)y^{\frac{1}{\delta}-\frac{1}{2}}$. Note that $g_{k,l}(1) = 0$, so that $f'_{k,l}(\frac{k}{k+l}) = 0$. Observe that y is a decreasing function of x. Thus, to show that $k/(k+l)$ is a local maximum of $f_{k,l}$, we need to show that $g'_{k,l}(1) > 0$. But

$$g'_{k,l}(1) = \left(\frac{2}{\delta} - 1\right)(\mu - \lambda) - (k+l)\mu\left(\frac{1}{\delta} - 1\right).$$

Thus,

$$g'_{k,l}(1) > 0 \iff \frac{k+l}{2} < \frac{\left(\frac{2}{\delta}-1\right)}{\frac{2}{\delta}-2}\left(1 - \frac{\lambda}{\mu}\right) \iff \frac{k+l}{2} < x_0,$$

with x_0 as in (3.76). If $m = \lfloor x_0 \rfloor$ or $n < 2x_0$, then $(k+l)/2 < x_0$, so that $g'_{k,l}(1) > 0$. This yields that $\frac{k}{k+l}$ is a local maximum of $f_{k,l}$. By Lemma 3.11.4, we deduce that the maximum of $f_{k,l}$ is achieved at either 0, $k/(k+l)$, or 1. Moreover, one can check that $f_{k,l}\left(\frac{k}{k+l}\right) = \varphi(k+l)$. Therefore,

$$\max_{[0,1]} f_{k,l} = \max\left(\varphi(k), \varphi(l), \varphi(k+l)\right).$$

Assume now $m = \lceil x_0 \rceil$ and $n \leq 2x_0$. If $(k,l) \neq (m-1, m)$, one can use the same arguments as above to identify the maximum of $f_{k,l}$. Thus, we are reduced to the case $k = m-1$, and $l = m$. As φ is increasing on $\{1, ..., m\}$, we have for any $x \in [0,1]$,

$$f_{m-1,m}(x) \leq \varphi(m)x^{2/\delta} + 2\mu(m(m-1))^{1-1/\delta}x^{1/\delta}(1-x)^{1/\delta} + \varphi(m)(1-x)^{2/\delta}.$$

As the function on the right-hand side, which we denote by f, is such that $f(x) = f(1-x)$ for any $x \in [0,1]$, we get by Lemma 3.11.4 that its maximum is achieved at 0 or 1/2. Thus,

$$\max_{x \in [0,1]} f_{m-1,m}(x) \leq \max\left(\varphi(m), f\left(\frac{1}{2}\right)\right).$$

We claim that $f(1/2) \leq \varphi(m)$, which amounts to say that

$$\left(1 - \frac{1}{m}\right)^{1-1/\delta} \leq \frac{1 - 2^{1-2/\delta}}{2^{1-2/\delta}}\left(1 - \frac{1}{m}\left(1 - \frac{\lambda}{\mu}\right)\right).$$

Note that as $m \geq x_0$, we have

$$1 - \frac{1}{m}\left(1 - \frac{\lambda}{\mu}\right) \geq \frac{1}{2/\delta - 1}.$$

Since $\delta \in (0,1)$ and $m \geq 2$, we only need to prove that

$$2^{-1+1/\delta} \leq \frac{1 - 2^{1-2/\delta}}{2^{1-2/\delta}} \frac{1}{2/\delta - 1},$$

which we can re-write as follow

$$2^{1/\delta} - 2^{1-1/\delta} - \frac{2}{\delta} + 1 \geq 0.$$

But the function on the left-hand side of the above inequality is increasing in $1/\delta$ on $[1,+\infty)$, and is equal to zero for $\delta = 1$. Thus, the above inequality is true for any $\delta \in (0,1)$, which proves our claim. We conclude that

$$\max_{x\in[0,1]} f_{m-1,m}(x) = \varphi(m).$$

We can deduce from (3.77) that

$$\max\left\{ \langle By, y\rangle : y \in [0,+\infty)^m, \sum_{i=1}^m y_i^\delta = 1 \right\} = \max_{1\le k\le m} \varphi(k).$$

$\square$

We come back now at the proof of case (c). As $\alpha \in (1,2)$, we have that $2(\alpha-1)/\alpha \in (0,1)$. From Lemma 3.11.3 and (3.72), we get

$$\begin{aligned} c &\ge \left\{ \max_{k\ge1} \left(\left(\frac{1}{b}\right)^{\frac{1}{\alpha-1}} + (k-1)\left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}} \right) k^{-(\alpha-1)} \right\}^{-(\alpha-1)} \\ &= \min_{k\ge1} \psi(k), \end{aligned}$$

where ψ is defined in the statement of Proposition 3.11.1. As we assume $1 \in \operatorname{supp}(\nu_1)\cap\operatorname{supp}(\nu_2)$, the matrix $P_k(p,q)$ defined in (3.63), is in the domain $\mathcal{D}$, and

$$W_\alpha\big(P_k(p,q)\big) = \psi(k),$$

which gives the first part of the claim in case (c).

Easy computations show that the function ψ defined in (3.64) is decreasing on $[0,t_0]$ and increasing on $[t_0,1]$, with

$$t_0 = \frac{1}{2-\alpha}\Big(1-\frac{q}{p}\Big).$$

Thus,

$$c = \min\left(\psi\left(\lfloor t_0\rfloor\right), \psi\left(\lceil t_0\rceil\right)\right).$$

(d). Let $1<\alpha<2$ and assume $1\in\operatorname{supp}(\nu_1)$, $\operatorname{supp}(\nu_2) = \{-1\}$ and $b > \frac{a}{2}$. Then,

$$c = \inf_{n\ge1} \inf\left\{ W_\alpha(A) : A \in \mathcal{H}_n^{(1)}, \lambda_A = 1, A_{i,j} \le 0, \forall i\neq j\right\},$$

Let $n\ge1$. We consider the minimization problem

$$\inf\left\{ W_\alpha(A) : A \in \mathcal{H}_n^{(1)}, \lambda_A = 1, A_{i,j} \le 0, \forall i\neq j\right\}.$$

The same argument as in case (c) justifies that the infimum is achieved at some $A\in\mathcal{H}_n^{(1)}$ for which 1 is a simple eigenvalue. The multipliers rule (see [40][Theorem 9.1]) asserts that there are some $(M,\gamma)\in\mathcal{H}_n^{(1)}\times\mathbb{R}$ such that $(M,\gamma)\neq(0,0)$, and

$$\forall i\neq j, M_{i,j}\ge 0, M_{i,j}A_{i,j} = 0, \text{ and } M_{i,i} = 0, \forall 1\le i\le n,$$

satisfying

$$\nabla W_\alpha(A) + M = \gamma x^t x,$$

where x is a unit eigenvector associated with the eigenvalue 1. We deduce that for any $i \neq j$,

$$\frac{a}{2}\alpha A_{i,j}\,|A_{i,j}|^{\alpha-2} + M_{i,j} = \gamma x_i x_j, \tag{3.80}$$

and for any $1 \leq i \leq n$,

$$b\alpha A_{i,i}\,|A_{i,i}|^{\alpha-2} = \gamma x_i^2. \tag{3.81}$$

The same argument as in case (c), shows that

$$\alpha W_\alpha(A) = \gamma. \tag{3.82}$$

Without loss of generality, we can assume x is of the form $x = (x_1, ..., x_k, x_{k+1}, ..., x_{k+l}, 0, ...0)$, with $x_1 > 0, ..., x_k > 0$, and $x_{k+1} < 0, ..., x_{k+l} < 0$.

Note that as $A_{i,j}M_{i,j} = 0$, $M_{i,j} \geq 0$, and $A_{i,j} \leq 0$, for any $i \neq j$, we get from (3.80), that for any $i \neq j$, $A_{i,j} \neq 0$ if and only if $x_i x_j < 0$. Thus, for all $i \neq j$, $A_{i,j} \neq 0$, if and only if (i,j) or $(j,i) \in \{1, ..., k\} \times \{k+1, ..., k+l\}$.

Let $(i,j) \in \{1, ..., k\} \times \{k+1, ..., k+l\}$. From (3.80), we have

$$A_{i,j} = -\left(\frac{2\gamma}{a\alpha}\right)^{\frac{1}{\alpha-1}} |x_i x_j|^{\frac{1}{\alpha-1}},$$

and for all $i \in \{1, ..., k+l\}$, we get by (3.81),

$$A_{i,i} = \left(\frac{\gamma}{b\alpha}\right)^{\frac{1}{\alpha-1}} |x_i|^{\frac{2}{\alpha-1}}.$$

The eigenvalue equation $Ax = x$, yields, for any $i \in \{1, ..., k\}$,

$$\left(\frac{\gamma}{b\alpha}\right)^{\frac{1}{\alpha-1}} |x_i|^{\frac{2}{\alpha-1}+1} + \left(\frac{2\gamma}{a\alpha}\right)^{\frac{1}{\alpha-1}} \sum_{k+1 \leq j \leq k+l} |x_i|^{\frac{1}{\alpha-1}} |x_j|^{\frac{1}{\alpha-1}+1} = |x_i|,$$

as $x_j < 0$ for $j \in \{k+1, ..., k+l\}$, and $x_i > 0$ for $i \in \{1, ..., k\}$.

Similarly, for any $i \in \{k+1, ..., k+l\}$,

$$-\left(\frac{\gamma}{b\alpha}\right)^{\frac{1}{\alpha-1}} |x_i|^{\frac{2}{\alpha-1}+1} - \left(\frac{2\gamma}{a\alpha}\right)^{\frac{1}{\alpha-1}} \sum_{1 \leq j \leq k} |x_i|^{\frac{1}{\alpha-1}} |x_j|^{\frac{1}{\alpha-1}+1} = -|x_i|.$$

Dividing in the two equations above by $|x_i|^{\frac{1}{\alpha-1}}$, we get

$$B^{(k,l)}y = \left(\frac{\gamma}{\alpha}\right)^{-\frac{1}{\alpha-1}} y^{-\frac{2-\alpha}{\alpha}}, \tag{3.83}$$

with $y \in \mathbb{R}^{k+l}$, such that $y_i = |x_i|^{\frac{1}{\alpha-1}+1}$, for all $i \in \{1, ..., k+l\}$, and

$$B^{(k,l)} = \left(\begin{array}{c|c} \left(\frac{1}{b}\right)^{\frac{1}{\alpha-1}} I_k & \left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}} U_{k,l} \\ \hline \left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}}\, {}^tU_{k,l} & \left(\frac{1}{b}\right)^{\frac{1}{\alpha-1}} I_l \end{array}\right) \in \mathcal{H}_{k+l}^{(1)},$$

where $U_{k,l}$ is the matrix of size $k \times l$ whose entries are all equal to 1. As x is a unit vector, we have $\sum_{i=1}^{k+l} y^{\frac{2(\alpha-1)}{\alpha}} = 1$. We deduce from (3.83), that

$$\left(\frac{\gamma}{\alpha}\right)^{-\frac{1}{\alpha-1}} = \langle B^{(k,l)}y, y\rangle.$$

Using (3.82) and the fact that A is a minimizer, we get

$$c = \left(\sup_{k,l\in\mathbb{N}} \sup \left\{ \langle B^{(k,l)}y, y \rangle : \sum_{i=1}^{k+l} y_i^{2(\alpha-1)/\alpha} = 1, y \in [0,+\infty)^{k+l} \right\} \right)^{-(\alpha-1)}. \quad (3.84)$$

In the following lemma, we compute the supremum of the left-hand side of the above inequality.

3.11.5 Lemma. *Let $\delta \in (0,1)$. Let $k, l \in \mathbb{N}$, such that $(k,l) \neq (0,0)$. We define for $\lambda, \mu \in \mathbb{R}_+$, the matrix*

$$B = \left(\begin{array}{c|c} \lambda I_k & \mu U_{k,l} \\ \hline \mu^t U_{k,l} & \lambda I_l \end{array} \right) \in \mathcal{H}_{k+l}^{(1)},$$

where $U_{k,l}$ is the matrix of size $k \times l$ whose entries are all equal to 1. We have,

$$\sup \left\{ \langle By, y \rangle : \sum_{i=1}^{k+l} y_i^{\delta} = 1, y \in [0,+\infty)^{k+l} \right\} = \max\left(\lambda, (\lambda+\mu)2^{1-2/\delta} \right).$$

Proof. With the same arguments as in the proof of Lemma 3.11.3, the supremum of the quadratic form defined by B on

$$\left\{ y \in [0,+\infty)^{k+l} : \sum_{i=1}^{k+l} y_i^{\delta} = 1 \right\},$$

is achieved at some y such that,

$$\forall i \in \{1,...,k+l\},\ y_i = s_i \mathbb{1}_{i \leq k'} + t_{k'+i} \mathbb{1}_{1 \leq i \leq l'},$$

with $s_1 > 0, ..., s_{k'} > 0$, and $t_{k'+1} > 0, ..., t_{k'+l'} > 0$, for some $k' \leq k$ and $l' \leq l$, such that the vector $z = (s_1, ..., s_{k'}, t_{k'+1}, ..., t_{k'+l'}) \in \mathbb{R}^{k'+l'}$, satisfies for some $\gamma \in \mathbb{R}$,

$$\tilde{B}z = \gamma z^{\delta-1},$$

where

$$\tilde{B} = \left(\begin{array}{c|c} \lambda I_{k'} & \mu U_{k',l'} \\ \hline \mu^t U_{k',l'} & \lambda I_{l'} \end{array} \right) \in \mathcal{H}_{k'+l'}^{(1)}.$$

Without loss of generality, we can assume $k, l \geq 1$. Comparing the i^{th} and j^{th} coordinates of Bz, for $1 \leq i, j \leq k'$, we get

$$\lambda(s_i - s_j) = \gamma\left(s_i^{\delta-1} - s_j^{\delta-1} \right).$$

If $\lambda = 0$, then it immediately yields $s_i = s_j$. If $\lambda \neq 0$, as $\delta \in (0,1)$, we see that if $s_i \neq s_j$, the terms on the left-hand side, and the right-hand side must have opposite signs. Thus $s_i = s_j$ for any $i, j \in \{1, ..., k'\}$. Similarly, comparing the i^{th} and j^{th} coordinates of By, for $i, j \in \{k'+1, ..., k'+l'\}$, yields that $t_i = t_j$, for all $i, j \in \{k'+1, ..., k'+l'\}$. We can write

$$\forall i \in \{1, ..., k'+l'\},\ z_i = s\mathbb{1}_{i \leq k'} + t\mathbb{1}_{k'+1 \leq i \leq k'+l'},$$

for some $s, t \in (0, +\infty)$. As $\sum_{i=1}^{k'+l'} z_i^\delta = 1$, we have $k's^\delta + l't^\delta = 1$. Let $v = (k'^{1/\delta}s, l'^{1/\delta}t)$. Then,

$$\langle \tilde{B}z, z\rangle = \lambda(k's^2 + l't^2) + 2\mu k'l'ts = \big\langle M(k', l')v, v\big\rangle,$$

where

$$M^{(k',l')} = \begin{pmatrix} \lambda k'^{1-2/\delta} & \mu(k'l')^{1-1/\delta} \\ \mu(k'l')^{1-1/\delta} & \lambda l'^{1-2/\delta} \end{pmatrix}.$$

Thus,

$$\begin{aligned} \sup\Big\{ \langle By, y\rangle : \sum_{i=1}^{k+l} y_i^\delta = 1, y \in [0, +\infty)^{k+l}\Big\} &= \sup_{\substack{1\le k'\le k \\ 1\le l'\le l}} \sup_{\substack{v=(s,t) \\ s^\delta+t^\delta=1, s,t\ge 0}} \langle M^{(k',l')}v, v\rangle \\ &= \sup_{\substack{v=(s,t) \\ s^\delta+t^\delta=1, s,t\ge 0}} \sup_{\substack{1\le k'\le k \\ 1\le l'\le l}} \langle M^{(k',l')}v, v\rangle. \end{aligned}$$

But for fixed $v \in \mathbb{R}^2$, as $\delta \in (0, 1)$, we easily see that the maximum of $\langle M^{(k',l')}v, v\rangle$ is achieved at $k' = l' = 1$. Thus,

$$\sup\Big\{ \langle By, y\rangle : \sum_{i=1}^{k+l} y_i^\delta = 1, y \in [0, +\infty)^{k+l}\Big\} = \sup_{\substack{v=(s,t) \\ s^\delta+t^\delta=1, s,t\ge 0}} \langle M^{(1,1)}v, v\rangle.$$

From Lemma 3.11.3, we get

$$\sup_{\substack{v=(s,t) \\ s^\delta+t^\delta=1, s,t\ge 0}} \langle M^{(1,1)}v, v\rangle = \max\Big(\lambda, (\lambda+\mu)2^{1-2/\delta}\Big),$$

which yields the claim. □

We come back now to the proof of case (d). By Lemma 3.11.5 and (3.84), we get

$$c = \max\Big(b, \frac{2}{(p+q)^{\alpha-1}}\Big),$$

which gives the claim.

(e). Let $1 < \alpha < 2$, and assume $1 \in \mathrm{supp}(\nu_2)$ and $\mathrm{supp}(\nu_1) = \{-1\}$. Then,

$$c \ge \inf_{n\ge 2} \inf\Big\{W_\alpha(A) : A \in \mathcal{H}_n^{(2)}, A_{i,i} \le 0, \forall i \in \mathbb{N}, \lambda_A = 1\Big\}.$$

Let $n \ge 2$. We consider the minimization problem

$$\inf\Big\{W_\alpha(A) : A \in \mathcal{H}_n^{(2)}, A_{i,i} \le 0, \forall i \in \mathbb{N}, \lambda_A = 1\Big\}.$$

Similar arguments as in case (c) and (d) show that the infimum is achieved at some A such that $A_{i,i} = 0$ for all $1 \le i \le n$. By the multipliers rule and Lemma 3.11.2, we deduce that for any $i \ne j$,

$$A_{i,j} = \left(\frac{2\gamma}{a\alpha}\right)^{\frac{1}{\alpha-1}} X_{i,j}\, |X_{i,j}|^{\frac{1}{\alpha-1}-1},$$

where $\gamma = \alpha W_\alpha(A)$, and $X \in \mathcal{H}_n^{(2)}$ is such that $0 \leq X \leq \mathbb{1}_{E_1(A)}$, and $\mathrm{tr}X = 1$. We deduce that $\mathrm{tr}AX = 1$. This yields,

$$\left(\frac{2\gamma}{a\alpha}\right)^{\frac{1}{\alpha-1}} \sum_{i \neq j} |X_{i,j}|^{\frac{1}{\alpha-1}+1} = 1.$$

As $W_\alpha(A) = \frac{\gamma}{\alpha}$, we have

$$W_\alpha(A) = \frac{a}{2}\Big(\sum_{i\neq j} |X_{i,j}|^{\frac{1}{\alpha-1}+1}\Big)^{-(\alpha-1)} \geq \frac{a}{2}\Big(\max_{\substack{\mathrm{tr}X=1\\ X\geq 0}} \sum_{i\neq j} |X_{i,j}|^{\frac{1}{\alpha-1}+1}\Big)^{-(\alpha-1)}. \quad (3.85)$$

In the following lemma, we compute the maximum on the right-hand side.

3.11.6 Lemma. *Let $\beta \geq 2$. We have for any $n \in \mathbb{N}$, $n \geq 2$,*

$$\max\Big\{\sum_{1\leq i\neq j\leq n} |X_{i,j}|^\beta : X \in \mathcal{H}_n^{(2)}, X \geq 0, \mathrm{tr}X = 1\Big\} = \max_{2\leq k\leq n} (k-1)k^{1-\beta}.$$

Proof. Let $\xi : X \in \mathcal{H}_n^{(2)} \mapsto \sum_{i\neq j} |X_{i,j}|^\beta$. Note that ξ is convex, and that the constraints set,

$$S = \Big\{X \in \mathcal{H}_n^{(2)} : X \geq 0, \mathrm{tr}X = 1\Big\},$$

is also convex. As a consequence of [85, Corollary 18.5.1], ξ attains its maximum at an extreme point of the set S, which is of the form xx^*, with x a unit vector of $\mathbb{C}^n$. We deduce that,

$$\max_S \xi = \max\Big\{\sum_{1\leq i\neq j\leq n} |x_i x_j|^\beta : x \in \mathbb{C}^n, ||x|| = 1\Big\}.$$

We can re-write the maximum on the right-and side of the above equation as,

$$\max\Big\{\langle By, y\rangle : \forall i \in \{1, ..., n\},\ y_i \geq 0, \sum_{i=1}^n y_i^{2/\beta} = 1\Big\},$$

where

$$B = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \in \mathcal{H}_n^{(1)}.$$

Applying the result of Lemma 3.11.3, with $\delta = 2/\beta$, we get the claim. $\square$

We come back at the proof of Proposition 3.11.1, (e). Note that, as $1 < \alpha < 2$, we have $1 + \frac{1}{\alpha-1} \geq 2$. From (3.85) together with Lemma 3.11.6, we get

$$c \geq \frac{a}{2}\Big(\max_{n\geq 2}(n-1)n^{-\frac{1}{\alpha-1}}\Big)^{-(\alpha-1)} = \frac{a}{2}\min \frac{n}{(n-1)^{\alpha-1}}.$$

But,

$$\frac{a}{2}\frac{n}{(n-1)^{\alpha-1}} = W_\alpha(P_n(0,1)),$$

where $P_n(0,1)$ is defined in (3.63). As $1 \in \operatorname{supp}(\nu_2)$, we have $P_n(0,1) \in \mathcal{D}$, which ends the proof of the case (e).

(f). Assume finally $1 < \alpha < 2$, and $\operatorname{supp}(\nu_1) = \operatorname{supp}(\nu_2) = \{-1\}$. Let $n \geq 1$ and consider the minimization problem

$$\inf \left\{ W_\alpha(A) : A \in \mathcal{H}_n^{(1)}, \lambda_A = 1, A_{i,j} \leq 0, \forall i \leq j \right\}.$$

The same arguments as in the case (e), show that the minimizer A is such that $A_{i,i} = 0$ for all $i \in \{1, ..., n\}$. If A is a simple eigenvalue of A, then, the same analysis can be carried as in the case (d), and yields

$$W_\alpha(A) \geq \left(\sup_{k,l \in \mathbb{N}} \sup \left\{ \langle By, y \rangle : \sum_{i=1}^{k+l} y_i^{2(\alpha-1)/\alpha} = 1, y \in [0, +\infty)^{k+l} \right\} \right)^{-(\alpha-1)},$$

with

$$B = \left(\begin{array}{c|c} 0_k & \left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}} U_{k,l} \\ \hline \left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}} {}^t U_{k,l} & 0_l \end{array} \right) \in \mathcal{H}_{k+l}^{(1)},$$

where $U_{k,l}$ is the matrix of size $k \times l$ whose entries are all equal to 1, and 0_k, 0_l are the null matrices of sizes $k \times k$ and $l \times l$ respectively. Due to Lemma 3.11.5, we have

$$\sup_{k,l \in \mathbb{N}} \sup \left\{ \langle By, y \rangle : \sum_{i=1}^{k+l} y_i^{2(\alpha-1)/\alpha} = 1, y \in [0, +\infty)^{k+l} \right\} = \left(\frac{2}{a}\right)^{\frac{1}{\alpha-1}} 2^{-\frac{1}{\alpha-1}}.$$

Therefore, $W_\alpha(A) \geq a$.

Now, if 1 is not a simple eigenvalue of A, then we have by [104, Theorem 3.32],

$$W_\alpha(A) = \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha = \frac{a}{2} \sum_{i,j} |A_{i,j}|^\alpha \geq \frac{a}{2} \sum_{i=1}^{n} |\lambda_i|^\alpha \geq a,$$

where $\lambda_1, ..., \lambda_n$ are the eigenvalues of A.

In both cases, $W_\alpha(A) \geq a$. We deduce that $c \geq a$, and as

$$W_\alpha \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = a,$$

we get the claim. □

4. Large deviations of traces of random matrices

This chapter is based on the article *On the large deviations of traces of random matrices*, arXiv:1605.03894, accepted for publication in the *Annales de l'Institut Henri Poincaré*.

4.1 Introduction

We are concerned in this chapter with another instance of heavy-tail phenomena occurring in the large deviations of random matrices, that is the large deviations of the traces of powers of random matrices (equivalently, the moments of the empirical spectral measure). Indeed, in most cases - with the notable exception of bounded entries - the moments of the spectral measure do not have exponential moments, and therefore one can expect a heavy-tail phenomenon to hold. The main feature of these large deviations we want to advertise in this chapter, is that although the traces of powers of random matrices are empirical mean of correlated random variables, the lack of exponential integrability enable us to get round the correlation in the spectrum and get full LDPs.

Note that since the map which associates to a probability measure on $\mathbb{R}$, its p^{th} moment is not continuous for the weak topology, one cannot derive, by a contraction principle, LDPs for the p^{th} moment of the empirical spectral measure, from the already known large deviations principles for the empirical spectral measure.

Moderate deviations of certain traces of convex perturbation of the GUE multi-matrix model have been investigated in [49]. In the case where the entries are not centered, some results of large deviations for the moments of the empirical spectral measure are known. In the case of symmetric Bernoulli matrices, we know by [45, Theorem 1.5] that the centered traces satisfy moderate deviations principles with an explicit rate function. A large deviations principle for the traces of Bernoulli matrices is derived in [38, Theorem 4.1], as a consequence of the LDP of Erdös-Renyi graphs with fixed parameter p, with respect to the cut metric, but only with "graphon" scaling.

4.2 Main results

Our aim goal in this chapter is to derive LDPs for the moments of the empirical (spectral) measure in three cases : the case of β-ensembles for convex potential with polynomial growth in section 4.4, the case of Gaussian Wigner matrices in section 4.3, and the case of Wigner matrices without Gaussian tails in section 4.5.

We start with the case of the β-ensembles. We recall that a β-ensemble associated with the potential V is defined by the probability measure,

$$\mathbb{P}^n_{V,\beta} = \frac{1}{Z^n_{V,\beta}} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i, \tag{4.1}$$

with

$$Z^n_{V,\beta} = \int \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i,$$

where we assume the following growth condition on V,

$$\liminf_{|x|\to+\infty} \frac{V(x)}{\beta' \log|x|} > 1, \tag{4.2}$$

for some $\beta' \geq \beta, \beta' > 1$. In the case V is a convex potential with polynomial growth, we have the following result.

4.2.1 Theorem. *Let $\alpha \geq 2$ and $\beta > 0$. Let*

$$\forall x \in \mathbb{R},\ V(x) = b|x|^\alpha + w(x), \tag{4.3}$$

where w is a continuous convex function such that $w(x) = o_{\pm\infty}(|x|^\alpha)$. Let $p \in \mathbb{N}$, $p > \alpha$. For any $\lambda_1, ..., \lambda_n \in \mathbb{R}^n$, we denote by $m_{p,n}$,

$$m_{p,n} = \frac{1}{n}\sum_{i=1}^n \lambda_i^p.$$

Under $\mathbb{P}^n_{V,\beta}$, the sequence $(m_{p,n})_{n\geq 1}$ satisfies a large deviations principle with speed $n^{1+\frac{\alpha}{p}}$ and good rate function J_p, where $\mathbb{P}^n_{V,\beta}$ is defined in (1.11). If p is even,

$$J_p(x) = \begin{cases} b\left(x - \langle \sigma^V_\beta, x^p\rangle\right)^{\frac{\alpha}{p}} & \text{if } x \geq \langle \sigma^V_\beta, x^p\rangle, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\langle \sigma^V_\beta, x^p\rangle$ denotes the p^{th} moment of the equilibrium measure of $\mathbb{P}^n_{V,\beta}$, and if p is odd, J_p is defined by,

$$\forall x \in \mathbb{R},\ J_p(x) = b\left|x - \langle \sigma^V_\beta, x^p\rangle\right|^{\frac{\alpha}{p}}.$$

4.2.2 Remark. The rate function in Theorem 4.2.1 is the same as the rate function of the LDP of

$$(\langle \sigma^V_\beta, x^p\rangle + \frac{1}{n}\sum_{i=1}^n X_i^p)_{n\in\mathbb{N}},$$

where $(X_i)_{i\geq 1}$ are i.i.d random variables with law $e^{-nV(x)}dx/Z_V$, where we denote $Z_V = \int e^{-nV(x)}dx$ (see Lemma 4.4.10). This indicates that the logarithmic interaction between the particles of the Coulomb gas becomes negligible when one is considering large deviations of $m_{p,n}$.

4.2.3 Remark. One can also derive a LDP of the even moments of the empirical measure, say $m_{2p,n}$, under $\langle \sigma_\beta^V, x^{2p}\rangle$, with speed n^2. Indeed, the proof of the large deviations of the empirical measure yields the asymptotics of the partition function $Z_{V,\beta}^n$ at the exponential scale n^2 (see [3, Theorem 2.6.1]). But the scaled logarithmic moment generating function of $m_{2p,n}$ at some $t<0$, is finite, and is actually equal to the ratio of partition functions $Z^n_{V-tx^{2p},\beta}/Z^n_{V,\beta}$. Gärtner-Ellis theorem (see [43, Theorem 2.3.6]), thus yields a LDP with speed n^2 of $\langle L_n, x^{2p}\rangle$ on $(-\infty, \langle \sigma_\beta^V, x^{2p}\rangle)$.

In the following we will denote by τ_n, the linear form $\frac{1}{n}\mathrm{tr}$ on $\mathcal{H}_n^{(\beta)}$. In the case of Gaussian Wigner matrices, we have the following result.

4.2.4 Theorem. *Let $p\in\mathbb{N}$, $p\geq 3$. Let X be a centered Wigner matrix with Gaussian entries, such that $\mathbb{E}|X_{1,2}|^2 = 1$. The sequence $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$, follows a LDP with speed $n^{1+\frac{2}{p}}$, and good rate function K_p. If p is even, K_p is given by,*

$$\forall x\in\mathbb{R},\ K_p(x) = \begin{cases} c\left(x - C_{p/2}\right)^{\frac{2}{p}} & \text{if } x\geq \langle \mu_{sc}, x^p\rangle,\\ +\infty & \text{if } x < \langle \mu_{sc}, x^p\rangle,\end{cases}$$

where $C_{p/2}$ is the $(p/2)^{th}$ Catalan number, whereas if p is odd,

$$K_p(x) = c|x|^{\frac{2}{p}}.$$

where

$$c = \inf\{q(H) : \mathrm{tr}H^p = 1, H\in\cup_m \mathcal{H}_m^{(\beta)}\},$$

with q is such that $Z^{-1}e^{-q}d\ell_n^{(\beta)}$ is the law of X, where $\ell_n^{(\beta)}$ is the Lebesgue measure on $\mathcal{H}_n^{(\beta)}$.

Moreover, in the case where the entries of X are real or $(\Re X_{1,2}, \Im X_{1,2})$ are independent with variance $1/2$, we can compute explicitly the constant c appearing in the above theorem.

4.2.5 Lemma. *Let X be a Gaussian Wigner satisfying the assumptions of Theorem 4.2.4. If X has real entries or $(\Re X_{1,2}, \Im X_{1,2})$ are independent with variance $1/2$, then*

$$c = \frac{1}{2}\min\Big(\frac{1}{\sigma^2}, \frac{1}{2\beta}\Big).$$

We consider now the so-called model of Wigner matrices without Gaussian tails defined in 3.2.1.

4.2.6 Theorem. *Let $p\in\mathbb{N}$, $p\geq 3$. Let X be a Wigner matrix without Gaussian tail. The sequence $(\tau_n(X/\sqrt{n})^p)_{n\geq 1}$ satisfies a large deviations principle with speed $n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$ and good rate function J_p. If p is even, J_p is given by*

$$\forall x\in\mathbb{R},\ K_{p,\alpha}(x) = \begin{cases} c_p\left(x - C_{p/2}\right)^{\alpha/p} & \text{if } x\geq C_{p/2},\\ +\infty & \text{otherwise},\end{cases}$$

where $C_{p/2}$ denotes the $\left(\frac{p}{2}\right)^{th}$ Catalan number, and if p is odd, the rate function J_p is given by

$$\forall x \in \mathbb{R}, \ K_{p,\alpha}(x) = c_p|x|^{\alpha/p},$$

where c_p is a constant depending on p, α, a and b.

Furthermore, if $\alpha \in (0,1]$ and p is even, then $c_p = \min\left(b, 2^{-\alpha/p}a\right)$.

4.2.7 Remark. Note that for $p = 2$, the trace of X^2 is a sum of i.i.d random variables, so that one can apply Cramer's theorem (see [43, Theorem 2.2.3]) in the case where the entries have finite Laplace transform, or Nagaev's truncation approach (see [80] or [53]) in the case where the entries have a tail distribution behaving as $e^{-ct^{\alpha}}$, with some $c > 0$, and $\alpha \in (0,2)$.

4.2.8 Remark. The constant c_p appearing in Theorem 4.2.6 is the solution of an optimization problem described in (4.35). We solve this optimization problem in section 4.5.10, in the easiest case when $\alpha \in (0,1]$ and p is even, and we give a lower a bound and upper bound in the case p is even and $\alpha \in (0,2)$.

4.3 The Gaussian case

We study in this section the question of the large deviations of the moments of the empirical spectral measure of a Wigner matrix with Gaussian entries. We will use an approach which is greatly inspired from Borell and Ledoux's proof of the LDP for Weiner chaos (see [32] [33], [69]).

As we will see in the proof, the deviations of the trace are created by translations of X of the form $n^{\frac{1}{2}+\frac{1}{p}}H$, where H is with bounded Hilbert-Schmidt norm. One of the central argument relies on the following lemma.

4.3.1 Lemma. *Let $\beta \in \{1,2\}$. We denote by $\mathcal{H}_n^{(\beta)}$ the set of symmetric matrices of size n when $\beta = 1$, and Hermitian matrices when $\beta = 2$. Let $|| \ ||_2$ denote the Hilbert-Schmidt norm on $\mathcal{H}_n^{(\beta)}$. Let X be a Wigner matrix whose entries are centered and have finite $(p+1)^{th}$ moment. For any $r > 0$,*

$$\sup_{\substack{||H||_2 \le r \\ H \in \mathcal{H}_n^{(\beta)}}} \left| \tau_n\left(\frac{X}{\sqrt{n}} + n^{1/p}H\right)^p - \langle \mu_{sc}, x^p\rangle - \mathrm{tr}H^p \right| \underset{N\to+\infty}{\longrightarrow} 0.$$

Proof. By Wigner's theorem (see [3, Lemmas 2.1.6, 2.1.7]) and Jensen's inequality, we only have to prove that for any $Y, H \in \mathcal{H}_n^{(\beta)}$,

$$|\mathrm{tr}(Y+H)^p - \mathrm{tr}Y^p - \mathrm{tr}H^p| \le 2^p \max_{1\le k\le p-1}\left\{\left(\mathrm{tr}|Y|^{p+1}\right)^{\frac{k}{p+1}}(\mathrm{tr}H^2)^{\frac{p-k}{2}}\right\}.$$

Let $Y, H \in \mathcal{H}_N^{(\beta)}$. Expanding the trace, and using the cyclicity of the trace, it suffices to prove that for any $s \in \{1,...,p\}$, $n_1,...,n_s \in \mathbb{N}$, $m_1,...,m_s \in \mathbb{N}$, such that $\sum_{i=1}^s n_i + \sum_{j=1}^s m_j = p$, we have

$$\left|\mathrm{tr}\left(Y^{n_1}H^{m_1}...Y^{n_s}H^{m_s}\right)\right| \le \left(\mathrm{tr}|Y|^{p+1}\right)^{\frac{k}{p+1}}\left(\mathrm{tr}|H|^2\right)^{\frac{p-k}{2}},$$

with $k = \sum_{i=1}^{s} n_i$. Applying Hölder's non-commutative inequality (see [22, Corollary IV.2.6]) with the exponents $\frac{p+1}{n_1}, \alpha, \frac{p+1}{n_2}, ..., \alpha, \frac{p+1}{n_s}$, with α such that

$$\frac{s}{\alpha} = 1 - \sum_{i=1}^{s} \frac{n_i}{p+1}, \tag{4.4}$$

we get,

$$\left|\mathrm{tr}\,(Y^{n_1} H^{m_1} ... Y^{n_s} H^{m_s})\,\right| \leq \left(\mathrm{tr}|Y|^{p+1}\right)^{\frac{1}{p+1}\sum_{i=1}^{s} n_i} \prod_{j=1}^{s} (\mathrm{tr}\,|H|^{\alpha m_i})^{\frac{1}{\alpha}}.$$

Note that when $s \geq 2$, we have from (4.4), $\alpha \geq 2$. If $s = 1$ and $m_1 = 1$, then as $p \geq 3$, (4.4) yields $\alpha m_1 \geq 2$. In any cases, $\alpha m_i \geq 2$ for any $i \in \{1, ..., s\}$. Therefore, for all $i \in \{1, ..., s\}$,

$$\mathrm{tr}\,|H|^{\alpha m_i} \leq \left(\mathrm{tr} H^2\right)^{\frac{\alpha m_i}{2}}.$$

Thus,

$$\left|\mathrm{tr}\,(Y^{n_1} H^{m_1} ... Y^{n_s} H^{m_s})\,\right| \leq \left(\mathrm{tr}|Y|^{p+1}\right)^{\frac{1}{p+1}\sum_{i=1}^{s} n_i} \left(\mathrm{tr} H^2\right)^{\frac{1}{2}\sum_{i=1}^{s} m_i},$$

which gives the claim. □

We are now ready to give a proof of Theorem 4.2.4. As in the proof of the LDP of Wiener chaoses (see [69] or [44, Theorem 5.1]), the proof of the upper bound relies on a reformulation of the deviations of the trace in terms of an enlargement of a properly chosen event. Then, the Gaussian isoperimetric inequality allows us to estimate the probability of such enlargement. Similarly as in Borell's proof of the lower bound, we use here the formula for the translation of the Gaussian measure.

Proof of Theorem 4.2.4. We closely follow the outline of proof of the large deviations of Wiener chaoses in [44, Section 5, Theorem 5.1]. Define for any $H \in \mathcal{H}_n^{(\beta)}$,

$$\varphi(H) = \langle \mu_{sc}, x^p \rangle + \mathrm{tr} H^p.$$

By homogeneity, we get that for all $s \in \mathbb{R}$,

$$K_p(s) = \inf \left\{ q(H) : s = \varphi(H), H \in \cup_{n \geq 1} \mathcal{H}_n^{(\beta)} \right\}.$$

Upper bound Let A be a closed subset of $\mathbb{R}$. We can assume without loss of generality that $\inf_A K_p > 0$, otherwise there is nothing to prove. Let $0 < r < \inf_A K_p$. As K_p is a good rate function, $\{s : K_p(s) \leq r\}$, is a compact subset which does not intersect A. Thus we can find some $\delta > 0$ such that

$$A \cap V_\delta(\{s : K_p(s) \leq r\}) = \emptyset,$$

where V_δ denote the δ-neighborhood. We define for any $n \in \mathbb{N}$,

$$\mathcal{K}_n = \left\{ H \in \mathcal{H}_n^{(\beta)} : q(H) \leq 1 \right\}.$$

Note that $\varphi(\sqrt{r}\mathcal{K}_n) \subset \{s : K_p(s) \leq r\}$. Thus, We deduce that

$$\mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in A\Big) \leq \mathbb{P}\Big(\tau_n\left(\frac{X}{\sqrt{n}}\right)^p \notin \varphi\Big(\sqrt{2r}\mathcal{K}_n\Big) + B(0,\delta)\Big).$$

Let

$$V = \Big\{Y \in \mathcal{H}_n^{(\beta)} : \sup_{H \in \mathcal{K}_n} \Big|\tau_n\Big(\frac{Y}{\sqrt{n}} + n^{1/p}H\Big)^p - \varphi(H)\Big| < \delta\Big\}.$$

Then,

$$\mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \notin \varphi(\sqrt{r}\mathcal{K}_n) + B(0,\delta)\Big) \leq \mathbb{P}\big(X \notin V + \sqrt{r}n^{\frac{1}{2}+\frac{1}{p}}\mathcal{K}_n\big).$$

By Lemma 4.3.1, we know that for n large enough, $\mathbb{P}(X \in V) \geq 1/2$. The Gaussian isoperimetric inequality (see [44, Theorem 4.3]) yields

$$\mathbb{P}\Big(X \notin V + \sqrt{r}n^{\frac{1}{2}+\frac{1}{p}}\mathcal{K}_n\Big) \leq e^{-rn^{1+\frac{2}{p}}}.$$

Therefore,

$$\mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in A\Big) \leq e^{-rn^{1+\frac{2}{p}}}.$$

Thus,

$$\limsup_{n\to+\infty} \frac{1}{n^{1+\frac{2}{p}}} \log \mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in A\Big) \leq -r.$$

Since the previous inequality is valid for any $0 < r < \inf_A J_p$, this yields the upper bound of the LDP.

Lower bound Let A be an open subset of $\mathbb{R}$. Let $s \in A$. There is some $\delta > 0$ such that $B(s,\delta) \subset A$. We can assume without loss of generality that $K_p(s) < +\infty$. Define for any $N \in \mathbb{N}$,

$$\forall t \in \mathbb{R},\ K_{p,n}(t) = \inf_{H \in \mathcal{H}_n^{(\beta)}} \{q(H) : t = \varphi(H)\}.$$

As $K_p = \inf K_{p,n}$, we deduce that there is some $r > 0$, such that for any $n \in \mathbb{N}$,

$$K_{p,n}(s) = \inf_{H \in r\mathcal{K}_n} \Big\{q(H) : s = \varphi(H)\Big\}.$$

Let $H \in r\mathcal{K}_n$ be such that $s = \langle \mu_{sc}, x^p\rangle + \mathrm{tr}H^p$. Then,

$$\mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in A\Big) \geq \mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in B(s,\eta)\Big) = \mathbb{P}(X \in V), \tag{4.5}$$

where

$$V = \{Y \in \mathcal{H}_n^{(\beta)} : \Big|\tau_n\Big(\frac{Y}{\sqrt{n}}\Big)^p - \varphi(H)\Big| < \eta\},$$

By Lemma 4.3.1, we know that $\mathbb{P}(X \in V - n^{\frac{1}{2}+\frac{1}{p}}H)$ goes to 1 as n goes to $+\infty$. But,

$$\begin{aligned}\mathbb{P}(X \in V) &= \mathbb{P}\big(X - n^{\frac{1}{2}+\frac{1}{p}}H \in V - n^{\frac{1}{2}+\frac{1}{p}}H\big)\\ &= \frac{1}{Z_n^{(\beta)}} \int_{V - n^{\frac{1}{2}+\frac{1}{p}}H} e^{-q(Y + n^{\frac{1}{2}+\frac{1}{p}}H)} d\ell_n^{(\beta)}(Y),\end{aligned}$$

where $d\ell_n^{(\beta)}$ denotes the Lebesgue measure on $\mathcal{H}_n^{(\beta)}$, and $Z_n^{(\beta)} = \int e^{-q(Y)} d\ell_n^{(\beta)}(Y)$. We re-write this probability as,

$$\mathbb{P}\left(X \in V\right) = e^{-q(H)n^{1+\frac{2}{p}}} \mathbb{E}\Big(\mathbb{1}_{\{X \in V - n^{\frac{1}{2}+\frac{1}{p}} H\}} e^{-n^{\frac{1}{2}+\frac{1}{p}} \Re\psi(H,Y)}\Big),$$

where ψ is the bilinear (or sesquilinear form if $\beta = 2$) form associated to the quadratic form q. Using Jensen's inequality, we get

$$\begin{aligned}\mathbb{P}\left(X \in V\right) &\geq e^{-q(H)n^{1+\frac{2}{p}}} \mathbb{P}\big(X \in V - n^{\frac{1}{2}+\frac{1}{p}} H\big) \\ &\quad \times \exp\Big(- n^{\frac{1}{2}+\frac{1}{p}} \mathbb{E}\Big(\Re\psi(H,Y) \frac{\mathbb{1}_{\{X \in V - n^{\frac{1}{2}+\frac{1}{p}} H\}}}{\mathbb{P}(X \in V - n^{\frac{1}{2}+\frac{1}{p}} H)}\Big)\Big).\end{aligned}$$

Using Cauchy-Schwarz inequality yields,

$$\mathbb{E}\Big(- \Re\psi(H,X) \frac{\mathbb{1}_{\{X \in V - n^{\frac{1}{2}+\frac{1}{p}} H\}}}{\mathbb{P}\big(X \in V - n^{\frac{1}{2}+\frac{1}{p}} H\big)}\Big) \geq - \frac{1}{\mathbb{P}(X \in V - n^{\frac{1}{2}+\frac{1}{p}} H)} \big(\mathbb{E}(\Re\psi(X,H))^2\big)^{1/2}.$$

But, as X is a Gaussian matrix with density $Z^{-1}e^{-q}$ with respect to the Lebesgue measure, we have

$$\mathbb{E}(\Re\psi(X,H))^2 = q(H).$$

Thus for n large enough so that $\mathbb{P}(X \in V - n^{\frac{1}{2}+\frac{1}{p}} H) \geq 1/2$, we have

$$\mathbb{P}\left(X \in V\right) \geq \frac{1}{2} \exp\big(- q(H)n^{1+\frac{2}{p}} - 2q(H)^{1/2} n^{\frac{1}{2}+\frac{1}{p}}\big).$$

Since $H \in r\mathcal{K}_n$, we get

$$\mathbb{P}\left(X \in V\right) \geq \frac{1}{2} \exp\big(- q(H)n^{1+\frac{2}{p}} - 2rn^{\frac{1}{2}+\frac{1}{p}}\big).$$

As the above inequality is true for any $H \in r\mathcal{K}_n$ such that $s = \varphi(H)$, we have

$$\mathbb{P}\left(X \in V\right) \geq \frac{1}{2} \exp\Big(- K_{p,n}(s) n^{1+\frac{2}{p}} - 2rn^{\frac{1}{2}+\frac{1}{p}}\Big).$$

We deduce from (4.5) and the fact that $K_p = \inf K_{p,n}$,

$$\liminf_{n\to+\infty} \frac{1}{n^{1+\frac{2}{p}}} \log \mathbb{P}\Big(\tau_n\Big(\frac{X}{\sqrt{n}}\Big)^p \in A\Big) \geq -K_p(s).$$

□

We end this section with a proof of Lemma 4.2.5.

Proof of Lemma 4.2.5. Under the assumptions of Lemma 4.2.4, we have

$$\forall H \in \mathcal{H}_n^{(\beta)},\ q(H) = \frac{1}{2\sigma^2} \sum_i H_{i,i}^2 + \frac{\beta}{2} \sum_{i<j} |H_{i,j}|^2.$$

Thus, for any $H \in \mathcal{H}_n^{(\beta)}$, we have

$$q(H) \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right) \mathrm{tr} H^2.$$

As $p \geq 2$, we get

$$q(H) \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right) |\mathrm{tr} H^p|^{2/p}. \tag{4.6}$$

This yields for any $s \in \mathbb{R}$,

$$c_p \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right).$$

Observe that we always have $c_p \leq 1/(2\sigma^2)$ by taking a size one matrix. Define

$$H = \begin{pmatrix} 0 & \lambda & \cdots & \lambda \\ \lambda & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda \\ \lambda & \cdots & \lambda & 0 \end{pmatrix} \in \mathcal{H}_n^{(1)},$$

with $\lambda = \left(\frac{1}{(n-1)^p + (n-1)}\right)^{1/p}$. We have $\mathrm{tr} H^p = 1$, and

$$q(H) = \frac{n(n-1)\beta}{4}\left(\frac{1}{(n-1)^p + (n-1)}\right)^{2/p} \underset{n\to+\infty}{\longrightarrow} \frac{\beta}{4}.$$

This yields $c = \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right)$, which ends the proof. □

4.4 Large deviations of moments of the empirical measure of β-ensembles

We now give a proof of Theorem 4.2.1. In order to ease the notation, we will write $\mathbb{P}_V^n$ for $\mathbb{P}_{V,\beta}^n$, as well as Z_V^n instead of $Z_{V,\beta}^n$. We will denote by L_n the empirical measure of a vector $(x_1, ..., x_n) \in \mathbb{R}^n$, which is defined by,

$$L_n = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i}.$$

4.4.1 Deviations inequalities and convergence of the moments

The first step of the proof of Theorem 4.2.1 will be to show, under the mild assumption (4.2) the convergence in expectation, of the moments of the empirical measure towards the moments of the equilibrium measure σ_β^V. To do so, we will need a control on the tail probability of

$$\max_{1\leq i\leq n} |\lambda_i|,$$

under $\mathbb{P}_V^n$. To this end we prove a more general deviations inequality, which will be crucial later.

4.4.1 Proposition. *Let $n \in \mathbb{N}$, $n \geq 2$. Under assumption (4.2), there is a constant $M_0 > 0$, depending only on V and β, such that for any $M \geq M_0$ and $1 \leq k \leq n$,*

$$\mathbb{P}_V^n\left(L_n\left(I_M^c\right) \geq \frac{k}{n}\right) \leq \exp\left(-CknV_M\right),$$

where $I_M = [-M, M]$, C is a positive constant depending on V and β, and where $V_M = \inf_{|\lambda|\geq M} V(\lambda)$.

4.4.2 Remark. The proof of this proposition looks like the proof of the LDP for the largest particle of β-ensembles (see [3, Theorem 2.6.6]), but note that this inequality is not a consequence of the LDP for the largest particle as we allow k to take all the range of integers from 1 to n. One can consider this deviation inequality as interpolating deviations inequalities of the largest particle ($k = 1$) and the empirical measure of β-ensembles ($k = n$). Later we will use this inequality with k and M depending on n, this is why we take some care to prove here a *non-asymptotic* deviation inequality.

In order to prove this deviation inequality, we will need a rough control on the ratio of the partition functions Z_V^n and $Z^{n-k}_{\frac{nV}{n-k}}$. This is the object of the following lemma.

4.4.3 Lemma. *There are some constants c_1, c_2 depending on V and β, such that for any $n \in \mathbb{N}$, and $k \leq n$,*

$$c_1 nk \leq \log \frac{Z_V^n}{Z^{n-k}_{\frac{nV}{n-k}}} \leq c_2 nk,$$

where Z_V^n, and $Z^{n-k}_{\frac{nV}{n-k}}$ are defined in (4.1).

Proof. From the invariance under permutation of the coordinates of the measures $\mathbb{P}_V^n$ we have

$$\frac{Z_V^n}{Z^{n-k}_{\frac{nV}{n-k}}} = \frac{n!}{Z^{n-k}_{\frac{nV}{n-k}}} \int_{|\lambda_1|\geq \ldots \geq |\lambda_n|} e^{-n\sum_{i=1}^n V(\lambda_i)} \prod_{1\leq i<j\leq n} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n d\lambda_i.$$

Splitting the λ_i's between the k first largest in absolute value and the rest, and using again the invariance under permutation of the coordinates, we can bring out the measure $\mathbb{P}^{n-k}_{\frac{nV}{n-k}}$, which gives

$$\frac{Z_V^n}{Z^{n-k}_{\frac{nV}{n-k}}} = \frac{n!}{(n-k)!} \mathbb{E}^{n-k}_{\frac{nV}{n-k}} \Bigg(\int_{|\lambda_1|\geq \ldots \geq |\lambda_k|} e^{-n\sum_{i=1}^k V(\lambda_i)} \prod_{1\leq i<j\leq k} |\lambda_i - \lambda_j|^\beta$$
$$\times e^{\beta(n-k)\sum_{i=1}^k \langle L_{n-k}, \log|\lambda_i - .|\rangle} \mathbb{1}_{\mathrm{supp}(L_{n-k})\subset[-\lambda_k,\lambda_k]} \prod_{i=1}^k d\lambda_i \Bigg),$$

where $L_{n-k} = \frac{1}{n-k}\sum_{i=k+1}^n \delta_{\lambda_i}$. We re-write this equality as the following,

$$\frac{Z_V^n}{Z^{n-k}_{\frac{nV}{n-k}}} = \frac{n!}{(n-k)!} \mathbb{E}^{n-k}_{\frac{nV}{n-k}} \Bigg(\int_{|\lambda_1|\geq \ldots \geq |\lambda_k|} e^{-k^2 \int_{x\neq y} f(x,y) dL_k(x) dL_k(y)}$$
$$\times e^{-(n-k)\sum_{i=1}^k (V(\lambda_i) - \beta\langle L_{n-k}, \log|\lambda_i - .|\rangle)} \mathbb{1}_{\mathrm{supp}(L_{n-k})\subset[-\lambda_k,\lambda_k]} \prod_{i=1}^k e^{-V(\lambda_i)} d\lambda_i \Bigg),$$

with $L_k = \frac{1}{k}\sum_{i=1}^k \delta_{\lambda_i}$, and $f(x,y) = \frac{1}{2}V(x) + \frac{1}{2}V(y) - \frac{\beta}{2}\log|x-y|$. Note that from the assumption (4.2) on V, we have

$$c := \inf\{f(x,y) : x \neq y\} > -\infty, \; c' := \inf\{V(x) - \beta\log|x-y| : |y| \leq |x|\} > -\infty.$$

Thus,

$$\frac{Z_V^n}{Z_{\frac{nV}{n-k}}^{n-k}} \leq \binom{n}{k} e^{-k^2 c} e^{-(n-k)kc'} \left(\int e^{-V(x)} dx \right)^k .$$

As $\binom{n}{k} \leq n^k$, we get

$$\frac{Z_V^n}{Z_{\frac{nV}{n-k}}^{n-k}} \leq e^{c_2 nk},$$

with c_2 some constant depending on V and β.

For the lower bound, we write similarly as for the upper bound,

$$\begin{aligned} \log \frac{Z_V^n}{Z_{\frac{nV}{n-k}}^{n-k}} &= \log \mathbb{E}_{\frac{nV}{n-k}}^{n-k} \Big(\int e^{-(n-1)\sum_{i=1}^k V(\lambda_i)} \prod_{1 \leq i < j \leq k} |\lambda_i - \lambda_j|^\beta \\ &\quad \times e^{\beta(n-k)\sum_{i=1}^k \langle L_{n-k}, \log|\lambda_i - .|\rangle} \prod_{i=1}^k e^{-V(\lambda_i)} d\lambda_i \Big). \end{aligned}$$

Using twice Jensen's inequality, we get

$$\begin{aligned} \log \frac{Z_V^n}{Z_{\frac{nV}{n-k}}^{n-k}} &\geq \mathbb{E}_{\frac{nV}{n-k}}^{n-k} \Big(\log \int e^{-(n-1)\sum_{i=1}^k V(\lambda_i)} \prod_{1 \leq i < j \leq k} |\lambda_i - \lambda_j|^\beta \\ &\quad \times e^{\beta(n-k)\sum_{i=1}^k \langle L_{n-k}, \log|\lambda_i - .|\rangle} \prod_{i=1}^k \frac{e^{-V(\lambda_i)}}{\int e^{-V(x)} dx} d\lambda_i \Big) + k \log \left(\int e^{-V(x)} dx \right). \\ &\geq -(n-1)k \Big(\int V(\lambda) \frac{e^{-V(\lambda)} d\lambda}{\int e^{-V(x)} dx} \Big) + \frac{k(k-1)\beta}{2} \Big(\int \log|\lambda - \mu| \frac{e^{-V(\lambda) - V(\mu)} d\lambda d\mu}{\left(\int e^{-V(x)} dx\right)^2} \Big) \\ &\quad + \beta k(n-k) \mathbb{E}_{\frac{nV}{n-k}}^{n-k} \Big(\int \langle L_{n-k}, \log|\lambda - .|\rangle \frac{e^{-V(\lambda)} d\lambda}{\int e^{-V(x)} dx} \Big) + k \log \Big(\int e^{-V(x)} dx \Big). \end{aligned}$$

But for any $\mu \in \mathbb{R}$,

$$\begin{aligned} \int \log|\lambda - \mu| e^{-V(\lambda)} d\lambda &= \int_0^{+\infty} \log x \left(e^{-V(\mu+x)} + e^{-V(\mu-x)} \right) dx \\ &\geq \int_0^1 \log x \left(e^{-V(\mu+x)} + e^{-V(\mu-x)} \right) dx. \end{aligned}$$

As $\inf V < -\infty$, we have

$$\int \log|\lambda - \mu| e^{-V(\lambda)} d\lambda \geq 2e^{-\inf V} \int_0^1 \log(x) dx = -2e^{-\inf V}.$$

Thus,

$$\mathbb{E}_{\frac{nV}{n-k}}^{n-k} \left(\int \langle L_{n-k}, \log|\lambda - .|\rangle \frac{e^{-V(\lambda)} d\lambda}{\int e^{-V(x)} dx} \right) \geq -\frac{2e^{-\inf V}}{\int e^{-V(x)} dx}.$$

We can conclude that

$$\log \frac{Z_V^n}{Z_{\frac{nV}{n-k}}^{n-k}} \geq c_1 nk,$$

with c_1 a constant depending on V and β.

□

We are now ready to give a proof of Proposition 4.4.1.

Proof of Proposition 4.4.1. We can write as in the proof of Lemma 4.4.3,

$$\begin{aligned}\mathbb{P}_V^n\Big(L_n\left(I_M^c\right) \geq \frac{k}{n}\Big) &\leq \frac{n!}{(n-k)!}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\\ &\times \mathbb{E}^{n-k}_{\frac{nV}{n-k}}\Big(\int_{|\lambda_1|\geq\ldots\geq|\lambda_k|\geq M} e^{-n\sum_{i=1}^k V(\lambda_i)} \prod_{1\leq i<j\leq k}|\lambda_i-\lambda_j|^\beta\\ &\times e^{\beta(n-k)\sum_{i=1}^k\langle L_{n-k},\log|\lambda_i-.|\rangle}\mathbb{1}_{\mathrm{supp}(L_{n-k})\subset[-\lambda_k,\lambda_k]}\prod_{i=1}^k d\lambda_i\Big),\end{aligned}$$

with $L_{n-k}=\frac{1}{n-k}\sum_{i=k+1}^n \delta_{\lambda_i}$.

As for all $x,y\in\mathbb{R}$, $\log|x-y|\leq\log(1+|x|)+\log(1+|y|)$, and for any $|x|\leq|y|$, $\log|x-y|\leq\log 2+\log(1+|x|)$, we get

$$\begin{aligned}\mathbb{P}_V^n\Big(L_n\left(I_M^c\right) \geq \frac{k}{n}\Big) \leq \frac{n!}{(n-k)!}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}e^{k(n-k)\log 2}&\\ \times \mathbb{E}^{n-k}_{\frac{nV}{n-k}}\Big(\int_{|\lambda_1|\geq\ldots\geq|\lambda_k|\geq M} e^{-n\sum_{i=1}^k V(\lambda_i)}e^{\beta k\sum_{i=1}^k\log(1+|\lambda_i|)}&\\ \times e^{\beta(n-k)\sum_{i=1}^k\log(1+|\lambda_i|)}\mathbb{1}_{\mathrm{supp}(L_{n-k})\subset[-\lambda_k,\lambda_k]}\prod_{i=1}^k d\lambda_i\Big)&.\end{aligned}$$

From (4.2), we deduce that there is some $c_0>0$, such that for $|y|$ large enough,

$$V(y)-\beta\log\left(1+|y|\right)\geq c_0V(y).$$

Thus, for M large enough,

$$\begin{aligned}\mathbb{P}_V^n\Big(L_n\left(I_M^c\right) \geq \frac{k}{n}\Big) &\leq \frac{n!}{(n-k)!}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\int_{|\lambda_1|\geq\ldots\geq|\lambda_k|\geq M} e^{-c_0n\sum_{i=1}^k V(\lambda_i)}\prod_{i=1}^k d\lambda_i\\ &=\binom{n}{k}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\left(\int_{|\lambda|\geq M}e^{-c_0nV(\lambda)}d\lambda\right)^k.\end{aligned}$$

But,

$$\int_{|\lambda|\geq M}e^{-c_0nV(\lambda)}d\lambda\leq e^{-c_0(n-1)V_M}\int e^{-V(\lambda)}d\lambda\leq c_3e^{-\frac{c_0}{2}nV_M},$$

with some constant $c_3>0$, and where we used in the last inequality the fact that $n\geq 2$. We deduce from Proposition 4.4.3 that for M large enough,

$$\mathbb{P}_V^n\Big(L_n\left(I_M^c\right) \geq \frac{k}{n}\Big)\leq (c_3n)^k e^{knc_2}e^{-\frac{c_0}{2}knV_M}.$$

As $V_M\to+\infty$ as $M\to+\infty$, we can find some constants $M_0>0$, and $C>0$, depending on V and β, such that for any $M>M_0$,

$$\mathbb{P}_V^n\Big(L_n\left(I_M^c\right) \geq \frac{k}{n}\Big)\leq e^{-CknV_M}.$$

$\square$

As a consequence of the previous Proposition 4.4.1, we have the convergence of the expectation under $\mathbb{P}_V^n$, of the moments of the empirical measure, as stated in the next corollary.

4.4.4 Corollary. *Under assumption* (4.2)*, we have for any* $p \in \mathbb{N}$*,*

$$\mathbb{E}_V^n \langle L_n, x^p \rangle \underset{n\to+\infty}{\longrightarrow} \langle \sigma_\beta^V, x^p \rangle,$$

where $\mathbb{E}_V^n$ *denotes the expectation with respect to* $\mathbb{P}_V^n$*.*

Proof. Since $(L_n)_{n\geq 1}$ follows a LDP with speed n^2 (see [3, Theorem 2.6.1]), and rate function whose minimum is achieved at σ_β^V, we deduce that $(L_n)_{n\in\mathbb{N}}$ converges weakly in probability to σ_β^V under $\mathbb{P}_V^n$. Thus, it is enough to show that for any $k \in \mathbb{N}$,

$$\sup_{n\geq N_0} \mathbb{E}_V^n \langle L_n, |x|^k \rangle < +\infty,$$

for some $N_0 \geq 1$.

Let $k \in \mathbb{N}$. We have $\langle L_n, |x|^k \rangle \leq \max_{1\leq i\leq n} |\lambda_i|^k$. Besides, we know by Proposition 4.4.1 that

$$\mathbb{P}_V^n\Big(\max_{1\leq i\leq n} |\lambda_i| > M \Big) \leq e^{-CnV_M},$$

for any $M > M_0$, where C and M_0 are some positive constants. Thus,

$$\mathbb{E}_V^n \max_{1\leq i\leq n} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{k-1} e^{-CnV_x} dx.$$

By assumption we know that for $|x|$ large enough, $V_x \geq \beta' \log|x|$, with $\beta' > 1$, so that for M_0 large enough,

$$\mathbb{E}_V^n \max_{1\leq i\leq n} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{k-1} x^{-C\beta' n} dx.$$

We deduce that for $n \geq (k+1)/C\beta'$, and M_0 large enough,

$$\mathbb{E}_V^n \max_{1\leq i\leq n} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{-2} dx = M_0^k + \frac{k}{M_0}, \tag{4.7}$$

which yields the claim. □

4.4.2 An exponential equivalence

The goal of this section is to prove that the large deviations of $m_{p,n}$ are due to the deviations of the $\log n$ largest in absolute value λ_i's . More precisely, we will prove the following proposition.

4.4.5 Proposition. *For any* $p \in \mathbb{N}$*,* $p > \alpha$*, and* $\lambda_1, ..., \lambda_n \in \mathbb{R}$*, we denote by* $T_{p,n}$ *the truncated moment*

$$T_{p,n} = \frac{1}{n} \sum_{i=1}^{\log n} \lambda_i^{*p},$$

where $\lambda_1^*, ..., \lambda_n^*$ *is the rearrangement of the* λ_i*'s by decreasing absolute values. Under the notation and assumption of Theorem 4.2.1, we have for any* $t > 0$*,*

$$\lim_{n\to+\infty} \frac{1}{n^{1+\alpha/p}} \log \mathbb{P}_V^n \left(|m_{p,n} - \langle \sigma_\beta^V, x^p \rangle - T_{p,N}| > t \right) = -\infty.$$

As a consequence of Proposition 4.4.1 and Corollary 4.4.4, we have the following result.

4.4.6 Proposition. *Under assumption (4.2), we have*

$$\frac{1}{n}\mathbb{E}_V^n\Big(\sum_{i=\log n+1}^{n}\lambda_i^{*p}\Big)\underset{n\to+\infty}{\longrightarrow}\langle\sigma_\beta^V,x^p\rangle.$$

Proof. Due to Corollary 4.4.4, we only need to prove

$$\frac{1}{n}\sum_{i=1}^{\log n}\mathbb{E}_V^n\lambda_i^{*p}\underset{n\to+\infty}{\longrightarrow}0.$$

From (4.7) we have

$$\sup_{n\geq N_0}\mathbb{E}_V^n|\lambda_1^*|^p<+\infty, \tag{4.8}$$

with $N_0\in\mathbb{N}$. Thus for any $n\geq N_0$,

$$\Big|\frac{1}{n}\sum_{i=1}^{\log n}\mathbb{E}_V^n\lambda_i^{*p}\Big|\leq\frac{\log n}{n}\sup_{n\geq N_0}\mathbb{E}_V^n|\lambda_1^*|^p\underset{n\to+\infty}{\longrightarrow}0.$$

□

Due to the previous proposition, in order to prove Proposition 4.4.5, it suffices to show that

$$\frac{1}{n}\sum_{i=\log n+1}^{n}\lambda_i^{*p}$$

concentrates at a speed higher than $e^{-n^{1+\alpha/p}}$. To this end, we will use concentration inequalities for α-convex measures from [27]. This is the object of the following proposition.

4.4.7 Proposition. *Let $\alpha\geq 2$. Let $g:\mathbb{R}^n\to\mathbb{R}$ be a 1-Lipschitz function with respect to $||\ ||_{\ell^\alpha}$. Under the notation and assumption of Theorem 4.2.1, we have for every $t>0$,*

$$\mathbb{P}_V^n\left(g-\mathbb{E}_V^n g>t\right)\leq\exp\Big(-\frac{bnt^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}}\Big).$$

In particular, if $f:\mathbb{R}\to\mathbb{R}$ is a 1-Lipschitz function, and $l,m\in\{1,...,n\}$, $l\leq m$, then for any $t>0$,

$$\mathbb{P}_V^n\Big(\frac{1}{n}\sum_{i=l}^{m}f\left(\overline{\lambda_i}\right)-\frac{1}{n}\mathbb{E}_V^n\sum_{i=l}^{m}f\left(\overline{\lambda_i}\right)>t\Big)\leq\exp\Big(-\frac{bN^2t^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}}\Big),$$

where $\overline{\lambda_1},...,\overline{\lambda_n}$ is the rearrangement of the λ_i's in ascending order.

Proof. Let

$$\forall\lambda\in\mathbb{R}^n,\ \Phi(\lambda)=n\sum_{i=1}^{n}V(\lambda_i)-\frac{\beta}{2}\sum_{i\neq j}\log|\lambda_i-\lambda_j|\,.$$

We claim that Φ is α-convex with respect to the norm $|| \, ||_{\ell^\alpha}$ on $\mathbb{R}^n$, more precisely we will show that for all $\lambda, \mu \in \mathbb{R}^n$,

$$\Phi(\lambda) + \Phi(\mu) - 2\Phi\left(\frac{\lambda+\mu}{2}\right) \geq \frac{bn}{2^{\alpha-1}} ||\lambda - \mu||_{\ell^\alpha}^\alpha. \tag{4.9}$$

Note that for any $k, l \in \{1, ..., n\}$,

$$\mathrm{Hess}\Big(-\beta \sum_{i \neq j} \log |\lambda_i - \lambda_j| \Big)_{k,l} = \begin{cases} -(\lambda_k - \lambda_l)^{-2} & \text{if } k \neq l, \\ \sum_{j \neq k} (\lambda_j - \lambda_k)^{-2} & \text{if } k = l, \end{cases}$$

which defines a non-negative matrix since for any $x \in \mathbb{R}^n$,

$$\sum_{k \neq l} (\lambda_k - \lambda_l)^{-2} x_k^2 - \sum_{k \neq l} (\lambda_k - \lambda_l)^{-2} x_k x_l = \sum_{k < l} (\lambda_k - \lambda_l)^{-2} (x_k - x_l)^2 \geq 0.$$

As by assumption

$$\forall x \in \mathbb{R}, \ V(x) = b|x|^\alpha + w(x),$$

with w a convex function, we have, with the above observation, for any $\lambda, \mu \in \mathbb{R}^n$,

$$\Phi(\lambda) + \Phi(\mu) - 2\Phi\left(\frac{\lambda+\mu}{2}\right) \geq bn\Big(\sum_{i=1}^n \lambda_i^\alpha + \sum_{i=1}^n \mu_i^\alpha - 2\sum_{i=1}^n \left(\frac{\lambda_i + \mu_i}{2}\right)^\alpha \Big).$$

Since $\alpha \geq 2$, we have for any $x, y \in \mathbb{R}$,

$$\frac{1}{2}x^\alpha + \frac{1}{2}y^\alpha \geq \left(\frac{x+y}{2}\right)^\alpha + \left(\frac{x-y}{2}\right)^\alpha.$$

This yields the desired inequality (4.9).

We know, by [27, Corollary 4.1], that (4.9) entails that for any 1-Lipschitz function with respect to $|| \, ||_{\ell^\alpha}$, $g : \mathbb{R}^n \to \mathbb{R}$, and every $t > 0$,

$$\mathbb{P}_V^n \left(g - \mathbb{E}_V^n g > t \right) \leq \exp\Big(- \frac{bnt^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}} \Big). \tag{4.10}$$

Let now $f : \mathbb{R} \to \mathbb{R}$ be a 1-Lipschitz function, and $k, l \in \{1, ..., n\}$, $k \leq l$. We set

$$\forall \lambda \in \mathbb{R}^n, \ g(\lambda) = \frac{1}{n} \sum_{i=l}^m f(\overline{\lambda_i}),$$

For any $\lambda, \mu \in \mathbb{R}^n$, we have by Cauchy-Schwartz inequality

$$g(\lambda) - g(\mu) \leq \frac{1}{n} \sum_{i=l}^m \left| \overline{\lambda_i} - \overline{\mu_i} \right| \leq \frac{1}{n^{1/2}} \left(\sum_{i=l}^m |\overline{\lambda_i} - \overline{\mu_i}|^2 \right)^{1/2} \leq \frac{1}{n^{1/2}} \left(\sum_{i=1}^N |\lambda_i - \mu_i|^2 \right)^{1/2},$$

where we used in the last inequality Hardy-Littlewood-Polyá rearrangement inequality. Thus, by Hölder inequality

$$g(\lambda) - g(\mu) \leq n^{-\frac{1}{\alpha}} ||\lambda - \mu||_{\ell^\alpha}.$$

This shows that g is $n^{-\frac{1}{\alpha}}$-Lipschitz with respect to the norm $|| \, ||_{\ell^\alpha}$. Applying (4.10) to g gives the second inequality in the statement. $\square$

In the following proposition, we use the concentration inequalities of Proposition 4.4.7, together with a truncation procedure and the deviations estimate of Proposition 4.4.1, to prove that

$$\frac{1}{n}\sum_{i=\log n+1}^{n}\lambda_i^{*p},$$

is exponentially equivalent to its expectation with respect to $\mathbb{P}_V^n$. Combining this with the result of Proposition 4.4.6,

$$\frac{1}{n}\sum_{i=\log n+1}^{n}\mathbb{E}_V^n\lambda_i^{*p}\underset{n\to+\infty}{\longrightarrow}\langle\mu_{sc},x^p\rangle,$$

we will get Proposition 4.4.5.

4.4.8 Proposition. *For any $t>0$,*

$$\limsup_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\Big(\Big|\sum_{i=\log n+1}^{n}\lambda_i^{*p}-\sum_{i=\log n+1}^{n}\mathbb{E}_V^n\lambda_i^{*p}\Big|\geq tn\Big)=-\infty,$$

where $\lambda_1^*,\dots,\lambda_n^*$ denotes the rearrangement of the λ_i's by decreasing absolute values.

Proof. To ease the notation, we set $k=\log n$. The first part of the argument consists in choosing the proper truncation level with respect to our exponential scale $n^{1+\alpha/p}$. For any $M_0>0$, we denote by F_{M_0} the function

$$\forall x\in\mathbb{R},\ F_{M_0}(x)=\begin{cases}\mathrm{sg}(x)\,(|x|\wedge M_0)^p & \text{if } p \text{ is odd},\\ (|x|\wedge M_0)^p & \text{if } p \text{ is even}.\end{cases}.$$

Let

$$M_0=\frac{n^{\frac{1}{\alpha(p-1)}(1-\frac{\alpha}{p})}}{(\log n)^{\frac{1}{\alpha}}}.$$

Note that,

$$\begin{aligned}\Big|\frac{1}{n}\sum_{i=k+1}^{n}\mathbb{E}_V^n\left(\lambda_i^{*p}-F_{M_0}\left(\lambda_i^*\right)\right)\Big| &\leq\frac{1}{n}\sum_{i=k+1}^{n}\mathbb{E}_V^n\,|\lambda_i^*|^p\,\mathbb{1}_{|\lambda_i^*|\geq M_0}\\ &\leq\frac{1}{nM_0}\sum_{i=k+1}^{n}\mathbb{E}_V^n\,|\lambda_i^*|^{p+1}\\ &\leq\frac{n-k}{nM_0}\mathbb{E}_V^n\,|\lambda_1^*|^{p+1}\underset{n\to+\infty}{\longrightarrow}0,\end{aligned}$$

using (4.8), and the fact that as $p>\alpha$, $M_0\to+\infty$. Thus, it suffices to prove that for any $t>0$,

$$\limsup_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\Big(\Big|\sum_{i=k+1}^{n}\lambda_i^{*p}-\sum_{i=k+1}^{n}\mathbb{E}_V^nF_{M_0}\left(\lambda_i^*\right)\Big|\geq tn\Big)=-\infty.$$

Note that,

$$\sum_{i=k+1}^{N}F_{M_0}\left(\lambda_i^*\right)=\sum_{j=(k-l)+1}^{N-l}F_{M_0}\left(\overline{\lambda_i}\right),$$

where $l = \mathrm{Card}\{i \in \{1, ..., k\} : \lambda_i^* > 0\}$. Since the function F_{M_0} is pM_0^{p-1}-Lipschitz, we have using a union bound and Proposition 4.4.7, for any $t > 0$,

$$\mathbb{P}_V^n\Big(\Big|\sum_{i=k+1}^n F_{M_0}(\lambda_i^*) - \sum_{i=k+1}^n \mathbb{E}_V^n F_{M_0}(\lambda_i^*)\Big| > tn\Big) \leq 2k \exp\Big(-\frac{1}{c_\alpha p^\alpha} t^\alpha n^{1+\frac{\alpha}{p}} \log n\Big),$$

where c_α is some constant depending on α. We can write,

$$\begin{aligned}
&\mathbb{P}_V^n\Big(\Big|\sum_{i=k+1}^n \lambda_i^{*p} - \sum_{i=k+1}^n \mathbb{E}_V^n F_{M_0}(\lambda_i^*)\Big| > nt\Big) \\
&\leq \mathbb{P}_V^n\Big(\Big|\sum_{i=k+1}^n F_{M_0}(\lambda_i^*) - \sum_{i=k+1}^n \mathbb{E}_V^n F_{M_0}(\lambda_i^*)\Big| > tn/2\Big) \\
&+ \mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*|} > tn/2\Big).
\end{aligned}$$

We saw by the concentration inequality above, that the deviations of the truncated moments at the level M_0 around its mean are exponentially negligible at the scale $n^{1+\alpha/p}$. We need now to prove that the contributions in the deviations of the truncated moments of the λ_i's above the level M_0 are also negligible. To do so, we will truncate one more time at a level R, chosen so that the deviation bound of Proposition 4.4.1 gives the right exponential estimate.

From (4.3), we have for M large enough,

$$\inf_{|x| \geq M} V(x) \geq \frac{b}{2} M^\alpha.$$

Proposition 4.4.1 yields that there are some constants $M_0 > 0$, and $C > 0$, depending on V and β, such that for any $M > M_0$, and $k \in \{1, ..., n\}$,

$$\mathbb{P}_V^n\Big(L_n(I_M^c) \geq \frac{k}{n}\Big) \leq \exp\left(-CknM^\alpha\right). \tag{4.11}$$

Let $R = e^{-1}\frac{n^{1/p}}{(\log n)^{1/2\alpha}}$. We have, with the inequality above, for n large enough,

$$\begin{aligned}
\mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*|} > tn/2\Big) &\leq \mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p\, \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq R} > tn/2\Big) \\
&+ \mathbb{P}_V^n\Big(L_n(I_R^c) \geq \frac{k}{n}\Big),
\end{aligned} \tag{4.12}$$

where L_n denotes the empirical measure of the λ_i's, and where $I_R = [-R, R]$. From (4.11), we deduce that,

$$\mathbb{P}_V^n\Big(L(I_R^c) \geq \frac{k}{n}\Big) \leq \exp\Big(-Ce^{-\alpha}(\log n)^{1/2} n^{1+\frac{\alpha}{p}}\Big).$$

We are reduced to show that the event $\{\sum_{i=k+1}^n |\lambda_i^*|^p\, \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq R} > tn/2\}$ is exponentially negligible at the scale $n^{1+\alpha/p}$. To this end, we will slice up the set $\{\lambda \in \mathbb{R} : M_0 \leq |\lambda| \leq R\}$ into $\log\log n$ small intervals $\{\lambda \in \mathbb{R} : M_l \leq |\lambda| \leq M_{l+1}\}$ for which we will use the deviation bound (4.11). At each step, we choose the largest bound so that the event $\{\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_l \le |\lambda_i^*| \le M_{l+1}} > \frac{tn}{2}\}$ is exponentially negligible by (4.11). For any $m \geq 1$, we set

$$q_m = \left(1 - \frac{\alpha}{p}\right)\left(\frac{1}{p} + \frac{\alpha}{p^2} + \dots + \frac{\alpha^{m-1}}{p^m} + \frac{\alpha^{m-1}}{p^m(p-1)}\right),$$

and

$$M_m = \frac{n^{q_m}}{(\log n)^{1/\alpha}},$$

Observe that $q_m \underset{n\to+\infty}{\longrightarrow} \frac{1}{p}$, and

$$\frac{1}{p} - q_m = O\Big(\Big(\frac{\alpha}{p}\Big)^m\Big).$$

Let $m = \lfloor c \log\log n \rfloor$ with c such that $q_m \geq \frac{1}{p} - \frac{1}{\log n}$. With this choice, we have

$$M_m \geq R.$$

Thus, slicing up the set $\{\lambda \in \mathbb{R} : M_0 \leq |\lambda| \leq R\}$, we get

$$\begin{aligned}
\mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*| \le R} > \frac{tn}{2}\Big) &\leq \mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*| \le M_n} > \frac{tn}{2}\Big)\\
&\leq \mathbb{P}_V^n\left(\sum_{l=0}^{m-1}\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_l \le |\lambda_i^*| \le M_{l+1}} > \frac{tn}{2}\right)\\
&\leq \mathbb{P}_V^n\left(\sum_{l=0}^{m-1} M_{l+1}^p L_n\left(I_{M_l}^c\right) > \frac{t}{2}\right).
\end{aligned}$$

Finally, a union bound gives

$$\mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*| \le R_n} > \frac{tn}{2}\Big) \leq \sum_{l=0}^{m-1} \mathbb{P}_V^n\Big(L_n\left(I_{M_l}^c\right) > \frac{t}{2mM_{l+1}^p}\Big).$$

Using (4.11), we get n large enough, and for all $0 \leq l \leq m$,

$$\mathbb{P}_V^n\Big(L(I_{M_l}^c) > \frac{t}{2mM_{l+1}^p}\Big) \leq \exp\Big(-\frac{Ctn^2M_l^\alpha}{2mM_{l+1}^p}\Big) \leq \exp\Big(-\frac{Ctn^{2+\alpha q_l - pq_{l+1}}(\log n)^{\frac{p}{\alpha}-1}}{2c\log\log n}\Big),$$

since $m \leq c \log\log n$. But,

$$\begin{aligned}
\alpha q_l - pq_{l+1} &= \left(1 - \frac{\alpha}{p}\right)\left(\frac{\alpha}{p} + \frac{\alpha^2}{p^2} + \dots + \frac{\alpha^l}{p^l} + \frac{\alpha^l}{p^l(p-1)}\right)\\
&\quad - \left(1 - \frac{\alpha}{p}\right)\left(1 + \frac{\alpha}{p} + \frac{\alpha^2}{p^2} + \dots + \frac{\alpha^l}{p^l} + \frac{\alpha^l}{p^l(p-1)}\right)\\
&= -\left(1 - \frac{\alpha}{p}\right).
\end{aligned}$$

Therefore,

$$\mathbb{P}_V^n\Big(L(I_{M_l}^c) > \frac{t}{2mM_{l+1}^p}\Big) \leq \exp\Big(-\frac{Ctn^{1+\frac{\alpha}{p}}(\log n)^\kappa}{2c\log\log n}\Big),$$

where $\kappa > 0$ as $p > \alpha$. We can conclude that,

$$\mathbb{P}_V^n\Big(\sum_{i=k+1}^n |\lambda_i^*|^p \mathbb{1}_{M_0\leq|\lambda_i^*|\leq R_n} > tn/2\Big) \leq c\log\log n \exp\Big(-\frac{Ctn^{1+\frac{\alpha}{p}}(\log n)^\kappa}{2c\log\log n}\Big),$$

which ends the proof. □

4.4.3 Large deviations principle for the truncated moments

Since we know from Proposition 4.4.5, that $(m_{p,n})_{n\in\mathbb{N}}$ is exponentially equivalent to

$$\big(\langle\sigma_\beta^V, x^p\rangle + T_{p,n}\big)_{n\in\mathbb{N}},$$

we only need to derive a LDP for $(T_{p,n})_{n\in\mathbb{N}}$, in order to get the large deviations principle of $(m_{p,n})_{n\in\mathbb{N}}$ (see [43, Theorem 4.2.13]).

4.4.9 Proposition. *Under the assumption of Theorem 4.2.1 and the notation of Proposition 4.4.5, the sequence $(T_{p,n})_{n\in\mathbb{N}}$ follows a LDP under the law $\mathbb{P}_V^n$, with speed $n^{1+\frac{\alpha}{p}}$, and good rate function I_p. If p is odd, I_p is defined by*

$$\forall x\in\mathbb{R},\ I_p(x) = b|x|^{\alpha/p},$$

and if p is even,

$$\forall x\in\mathbb{R},\ I_p(x) = \begin{cases} bx^{\alpha/p} & \text{if } x\geq 0,\\ +\infty & \text{otherwise.}\end{cases}$$

Proof. To ease the notation, we set in the following $k = \log n$.
Exponential tightness. Let

$$\forall\lambda\in\mathbb{R}^n,\ g(\lambda) = \Big(\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p}.$$

For $\lambda\in\mathbb{R}^n$, we set $l = \mathrm{Card}\{i\in\{1,...,k\} : \lambda_i^* > 0\}$. We can write

$$g(\lambda) = \Big(\sum_{i=1}^{k-l} |\overline{\lambda_i}|^p + \sum_{i=n-l+1}^n \overline{\lambda_i}^p\Big)^{1/p},$$

where $\overline{\lambda_1},...,\overline{\lambda_n}$ is the rearrangement of the λ_i's in ascending order. When l is fixed, as $p\geq\alpha$, we see that g is 1-Lipschitz with the same argument as in the proof of Proposition 4.4.7. Using a union bound, we get by Proposition 4.4.7, for any $t>0$,

$$\mathbb{P}_V^n\Big(\Big(\frac{1}{n}\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p} - \mathbb{E}_V^n\Big(\frac{1}{n}\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p} > t\Big) \leq k\exp\Big(-\frac{bt^\alpha n^{1+\frac{\alpha}{p}}}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}}\Big).$$

Besides, by Jensen's inequality

$$\mathbb{E}_V^n\Big(\frac{1}{n}\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p} \leq \Big(\mathbb{E}_V^n\frac{1}{n}\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p} \leq \Big(\frac{k}{n}\mathbb{E}_V^n|\lambda_1^*|^p\Big)^{1/p}.$$

From (4.8), we deduce

$$\mathbb{E}_V^n\Big(\frac{k}{n}\sum_{i=1}^k |\lambda_i^*|^p\Big)^{1/p} \underset{n\to+\infty}{\longrightarrow} 0.$$

From the above concentration inequality, we see that $(T_{p,n})_{n\in\mathbb{N}}$ is exponentially tight.

Upper bound. Observe that we only have to show that for any $x>0$,

$$\limsup_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\left(T_{p,n}\geq x\right)\leq -I_p(x). \tag{4.13}$$

In the case where p is even, it is clear that (4.13), is sufficient. In the case p is odd, observe that $\tilde{V}(x)=V(-x)$ satisfies the assumptions of Theorem 4.2.1. Note also that for any $x>0$,

$$\mathbb{P}_V^n\left(T_{p,n}\leq -x\right)=\mathbb{P}_{\tilde{V}}^n\left(T_{p,n}\geq x\right).$$

Therefore, if (4.13) is proven, and if p odd, then we have also for any $x>0$,

$$\limsup_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\left(T_{p,n}\leq -x\right)\leq -I_p(-x).$$

We now prove (4.13). Since $(\frac{1}{N}\sum_{i=1}^k|\lambda_i^*|^p)_{n\in\mathbb{N}}$ is exponentially tight, we only need to show that for any $M>x>0$, we have

$$\limsup_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\left(T_{p,n}\geq x,|\lambda_1^*|^p\leq Mn\right)\leq -I_p(x).$$

Let $M>x>0$. Since the event $\{T_{p,n}\geq x,|\lambda_1^*|^p\leq Mn\}$ is invariant under permutation of the λ_i's, we have

$$\begin{aligned}
&\mathbb{P}_V^n\left(T_{p,n}\geq x,|\lambda_1^*|^p\leq Mn\right)\\
&=\frac{n!}{Z_V^n}\int_{\substack{\sum_{i=1}^k\lambda_i^p\geq nx\\ |\lambda_n|\leq\ldots\leq|\lambda_1|\leq(Mn)^{1/p}}} e^{-n\sum_{i=1}^n V(\lambda_i)}\prod_{i<j}|\lambda_i-\lambda_j|^\beta\prod_{i=1}^n d\lambda_i.
\end{aligned}$$

Bounding the interaction term involving the k largest in absolute value λ_i's, we get

$$\begin{aligned}
&\mathbb{P}_V^n\left(T_{p,n}\geq x,|\lambda_1^*|^p\leq Mn\right)\\
&\leq\frac{n!}{(n-k)!}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\Big(2(nM)^{1/p}\Big)^{\beta nk}\int_{\substack{\sum_{i=1}^k\lambda_i^p\geq nx\\ |\lambda_k|\leq\ldots\leq|\lambda_1|}} e^{-n\sum_{i=1}^k V(\lambda_i)}\prod_{i=1}^k d\lambda_i\\
&\leq\binom{n}{k}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\Big(2(nM)^{1/p}\Big)^{\beta nk}\int_{\sum_{i=1}^k\lambda_i^p\geq nx} e^{-n\sum_{i=1}^k V(\lambda_i)}\prod_{i=1}^k d\lambda_i\\
&=\binom{n}{k}\frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z_V^n}\Big(2(nM)^{1/p}\Big)^{\beta nk}\Big(\int e^{-nV(\lambda)}d\lambda\Big)^k\,\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^k X_i^p\geq x\Big),
\end{aligned}$$

where $X_1, ..., X_k$ are independent and identically distributed random variables with law $d\mu_V = e^{-nV(x)}\frac{dx}{Z_n}$, where $Z_n = \int e^{-nV(x)}dx$. As

$$\int e^{-nV(x)}dx = e^{O(n)}, \text{ and } \log \frac{Z^{n-k}_{\frac{nV}{n-k}}}{Z^n_V} = O\left(n \log n\right),$$

from Lemma 4.4.3 (recall that $k = \log n$), it only remains to show that

$$\limsup_{n\to+\infty} \frac{1}{n^{1+\alpha/p}} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n} X_i^p \geq x\Big) \leq -I_p(x).$$

This is the object of the following lemma.

4.4.10 Lemma. *Let $(X_j)_{j\geq 1}$ be a sequence of independent and identically distributed random variables with law $d\mu_V = e^{-nV(x)}\frac{dx}{Z_n}$, where $Z_n = \int e^{-nV(x)}dx$, with V as in (4.3). Let $p \in \mathbb{N}$, $p > \alpha$.*

For any $x > 0$,

$$\limsup_{n\to+\infty} \frac{1}{n^{1+\alpha/p}} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n} X_i^p \geq x\Big) \leq -I_p(x),$$

with I_p as in Proposition 4.4.9.

Proof. Let $x > 0$. Set $Y_i = n^{1/\alpha}X_i$ for all $i \in \{1, ..., \log n\}$. We have

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n} X_i^p \geq x\Big) \leq \mathbb{P}\Big(\sum_{i=1}^{\log n} Y_i^p \geq x n^{1+\frac{p}{\alpha}}\Big) \leq \mathbb{P}\Big(\sum_{i=1}^{\log n} |Y_i|^p \geq x n^{1+\frac{p}{\alpha}}\Big).$$

Let $0 < t < 1$. As $\alpha \leq p$, we have $\alpha t/p < 1$. Using the fact that $(x+y)^s \leq x^s + y^s$, for any $s \in (0,1)$, $x, y \in \mathbb{R}^+$,

$$\begin{aligned}\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n} X_i^p \geq x\Big) &\leq \mathbb{P}\Big(\Big(\sum_{i=1}^{\log n} |Y_i|^p\Big)^{\frac{\alpha t}{p}} \geq x^{\frac{\alpha t}{p}} n^{t(1+\frac{\alpha}{p})}\Big)\\ &\leq \mathbb{P}\Big(\sum_{i=1}^{\log n} |Y_i|^{\alpha t} \geq x^{\frac{\alpha t}{p}} n^{t(1+\frac{\alpha}{p})}\Big).\end{aligned}$$

By Chernoff's inequality we get,

$$\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n} X_i^p \geq x\Big) \leq e^{-bx^{\frac{\alpha t}{p}} n^{t(1+\frac{\alpha}{p})}} \Big(\mathbb{E}\big(e^{b|Y_1|^{\alpha t}}\big)\Big)^{\log n}. \tag{4.14}$$

As for any $x \in \mathbb{R}$, $V(x) = b|x|^\alpha + w(x)$,

$$\mathbb{E}\big(e^{b|Y_1|^{\alpha t}}\big) = \frac{1}{Z_n'} \int e^{-b(|x|^\alpha - |x|^{\alpha t}) - nw\left(\frac{x}{n^{1/\alpha}}\right)} dx,$$

with

$$Z_n' = \int e^{-nV\left(\frac{x}{n^{1/\alpha}}\right)} dx.$$

On one hand,

$$\int e^{-b(|x|^\alpha-|x|^{\alpha t})-nw\left(\frac{x}{n^{1/\alpha}}\right)}dx \leq 2e^{n\inf w}\int_0^{+\infty} e^{-b(x^\alpha-x^{\alpha t})}dx.$$

Note that as w is convex, $\inf w > -\infty$. On the other hand, $Z_n' = e^{O(n)}$. Therefore,

$$\mathbb{E}\big(e^{b|Y_1|^{\alpha t}}\big) \leq e^{o\left(\frac{n^{1+\alpha/p}}{\log n}\right)}\int_0^{+\infty} e^{-b(x^\alpha-x^{\alpha t})}dx.$$

As $x\mapsto x^{\alpha-1}-tx^{\alpha t-1}$ is non-decreasing on $[1,+\infty)$, we have,

$$\begin{aligned}\int_0^{+\infty} e^{-b\left(x^\alpha-x^{\alpha t}\right)}dx &\leq e^b + \frac{1}{\alpha(1-t)}\int_1^{+\infty}\left(\alpha x^{\alpha-1}-\alpha t x^{\alpha t-1}\right)e^{-b\left(x^\alpha-x^{\alpha t}\right)}dx\\ &= e^b+\frac{1}{b\alpha(1-t)}.\end{aligned}$$

Take $t=t_n=1-1/(\log n)^2$. Then,

$$\mathbb{E}\left(e^{b|Y_1|^{\alpha t_n}}\right) = e^{o\left(\frac{n^{1+\alpha/p}}{\log n}\right)}.$$

Together with the bound (4.14), we get

$$\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^{\log n}X_i^p \geq x\Big) \leq -bx^{\frac{\alpha t_n}{p}}n^{-(1-t_n)\left(1+\frac{\alpha}{p}\right)}+o(1). \tag{4.15}$$

Taking the limsup as n goes to $+\infty$ we get the claim. □

Lower bound. Let $x\in\mathbb{R}$. We want to show that

$$\liminf_{\delta\to 0}\liminf_{n\to+\infty}\frac{1}{n^{1+\alpha/p}}\log\mathbb{P}_V^n\left(T_{p,n}\in(x-\delta,x+\delta)\right)\geq -I_p(x). \tag{4.16}$$

As $T_{p,n}$ converges to 0 in almost surely, it is enough to prove this bound for $x\neq 0$. With the same argument as for the upper bound, it suffices actually prove to the bound above only for $x>0$.

Let $x>0$ and $\delta>0$. We have for n large enough,

$$\mathbb{P}_V^n\left(T_{p,n}\in(x-\delta,x+\delta)\right)\geq \mathbb{P}_{V,\beta}^N\Big(\frac{1}{n}\lambda_1^{*p}\in\left(x-\delta/2,x+\delta/2\right),\forall i>1,|\lambda_i^*|\leq M\Big),$$

with $M>0$. By continuity, there is some $\varepsilon>0$ such that

$$\mathbb{P}_V^n\left(T_{p,n}\in(x-\delta,x+\delta)\right)\geq \mathbb{P}_V^n\Big(\frac{1}{n^{1/p}}\lambda_1^*\in(x^{1/p}-\varepsilon,x^{1/p}+\varepsilon),\forall i>1,|\lambda_i^*|\leq M\Big).$$

We have

$$\begin{aligned}&\mathbb{P}_V^n\left(T_{p,n}\in(x-\delta,x+\delta)\right)\\ &\geq n!\frac{Z_{\frac{nV}{n-1}}^{n-1}}{Z_V^n}\int_{\left|\frac{\lambda_1}{n^{1/p}}-x^{1/p}\right|<\varepsilon}d\lambda_1 e^{-nV(\lambda_1)}\mathbb{E}_{\frac{nV}{n-1}}^{n-1}\Big(\mathbb{1}_{L_{n-1}\in\mathcal{E}_M}e^{\beta(n-1)\langle\log(\lambda_1-.),L_{n-1}\rangle}\Big),\end{aligned}$$

where $L_{n-1} = \frac{1}{n-1}\sum_{i=2}^{n}\delta_{\lambda_i}$, and $\mathcal{E}_M = \{\mu \in \mathcal{P}(\mathbb{R}) : \mathrm{supp}(\mu) \subset [-M, M]\}$.Thus,

$$\mathbb{P}_V^n\left(T_{p,n} \in (x-\delta, x+\delta)\right) \geq n! \frac{Z^{n-1}_{\frac{nV}{n-1}}}{Z_V^n} \int_{\left|\frac{\lambda}{n^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-nV(\lambda)} d\lambda$$
$$\times e^{\beta(n-1)\log(n^{\frac{1}{p}}x^{\frac{1}{p}} - M - \varepsilon)} \mathbb{P}^{n-1}_{\frac{nV}{n-1}}\left(L_{n-1} \in \mathcal{E}_M\right).$$

As $w(y) = o_{\pm\infty}(|y|^{\alpha})$, we have

$$\int_{\left|\frac{\lambda}{n^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-nV(\lambda)} d\lambda \geq \int_{\left|\frac{\lambda}{n^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-(b+o(1))n\lambda^{\alpha}} dy$$
$$= e^{-(b-o(\varepsilon))n^{1+\frac{\alpha}{p}}x^{\frac{\alpha}{p}}} e^{o(n^{1+\frac{\alpha}{p}})}.$$

Thus,

$$\mathbb{P}_V^n\left(T_{p,n} \in (x-\delta, x+\delta)\right) \geq \frac{Z^{n-1}_{\frac{nV}{n-1}}}{Z_V^n} \mathbb{P}^{n-1}_{\frac{nV}{n-1}}\left(L_{n-1} \in \mathcal{E}_M\right) e^{-(b-o(\varepsilon))N^{1+\frac{\alpha}{p}}x^{\frac{\alpha}{p}}} e^{o(n^{1+\frac{\alpha}{p}})}.$$

But from Lemma 4.4.3 we know that $\log \frac{Z^{n-1}_{\frac{nV}{n-1}}}{Z_V^n} = O(n)$. Besides, by Proposition 4.4.1 (with $k = 1$), we have for M large enough,

$$\mathbb{P}^{n-1}_{\frac{nV}{n-1}}\left(L_{n-1} \in \mathcal{E}_M\right) \underset{n\to+\infty}{\longrightarrow} 1.$$

This concludes the proof of the lower bound (4.16). □

4.5 The case of Wigner matrices without Gaussian tails

We revisit in this section the model of Wigner matrices without Gaussian tails. We will show that the deviations of traces of powers of these matrices are due to a small proportion of the entries making deviations of order $n^{\frac{1}{2}+\frac{1}{p}}$. We start by a heuristic argument to give a idea of the nature of the deviations of the moments, and of the speed of the deviations.

4.5.1 Heuristics

We show here how one can get the lower bound of the LDP without much effort. The main fact which makes the argument work is the following : if we add to a given Hermitian matrix a low rank Hermitian matrix with not too large operator norm, then the map $A \mapsto \tau_n(A^p)$ is almost linear, where τ_n denotes the normalized trace $\frac{1}{n}\mathrm{tr}$ on $\mathcal{H}_n^{(\beta)}$. More precisely, we have the following lemma, whose proof is postponed at section 4.5.8.

4.5.1 Lemma. *Let $p \geq 2$. Let A and C be two Hermitian matrices of size N. Assume that C is of rank at most r. We have*

$$|\mathrm{tr}\,(A+C)^p - \mathrm{tr}A^p - \mathrm{tr}C^p| \leq 2^p r \max_{1\leq k\leq p-1} ||A||^k ||C||^{p-k},$$

where $||\ ||$ denotes the operator norm.

To make the argument clearer, let us assume X has entries distributed according to the exponential law with parameter b. We restrict ourself to the case where p is even. Let $\delta > 0$ and $\theta = (n\delta)^{1/p}$. Denoting $X^{(1)} = X - X_{1,1}e_1e_1^*$, where e_1 is the first coordinate vector of $\mathbb{C}^n$, we have

$$\begin{aligned}\mathbb{P}\left(\tau_n(X/\sqrt{n})^p \simeq C_{p/2}+\delta\right) &\gtrsim \mathbb{P}\Big(\tau_n\big(X^{(1)}/\sqrt{n}+\theta e_1e_1^*\big)^p \simeq C_{p/2}+\delta,\ \frac{X_{1,1}}{\sqrt{n}} \simeq \theta\Big)\\ &\gtrsim \mathbb{P}\Big(\tau_n\big(X^{(1)}/\sqrt{n}+\theta e_1e_1^*\big)^p \simeq C_{p/2}+\delta, ||X^{(1)}/\sqrt{n}|| \leq c\Big)\\ &\times \mathbb{P}\Big(\frac{X_{1,1}}{\sqrt{n}} \simeq \theta\Big),\end{aligned}$$

with some $c > 2$. As $||(X^{(1)}-X)/\sqrt{n}|| \to 0$ in probability, and

$$||X/\sqrt{n}|| \underset{n\to+\infty}{\longrightarrow} 2,$$

in probability by [8, Theorem 5.1] (or [3, Theorem 2.1.22, Exercise 2.1.27]), we have

$$\mathbb{P}\big(||X^{(1)}/\sqrt{n}|| \leq c\big) \underset{n\to+\infty}{\longrightarrow} 1.$$

By Lemma 4.5.1, we have

$$\begin{aligned}\mathbb{P}\Big(\tau_n(X/\sqrt{n})^p \simeq C_{p/2}+\delta\Big) &\gtrsim \mathbb{P}\Big(\tau_n\big(X^{(1)}/\sqrt{n}\big)^p + \frac{1}{n}\theta^p \simeq C_{p/2}+\delta, ||X^{(1,1)}/\sqrt{n}|| \leq c\Big)\\ &\times \mathbb{P}\Big(\frac{X_{1,1}}{\sqrt{n}} \simeq \theta\Big)\\ &\gtrsim \mathbb{P}\Big(\tau_n\big(X^{(1,1)}/\sqrt{n}\big)^p \simeq C_{p/2}, ||X^{(1)}/\sqrt{n}|| \leq c\Big)\\ &\times \mathbb{P}\Big(\frac{X_{1,1}}{\sqrt{n}} \simeq \theta\Big).\end{aligned}$$

Since $X_{1,1}$ has exponential law with parameter 1, we have

$$\mathbb{P}\Big(\frac{X_{1,1}}{\sqrt{n}} \simeq (n\delta)^{\frac{1}{p}}\Big) \simeq \exp\big(-b^{\frac{1}{2}+\frac{1}{p}}\delta\big).$$

But $\tau_n(X^{(1)}/\sqrt{n})^p$ converges to $C_{p/2}$ in probability, by Wigner's theorem (see [3, Lemmas 2.1.6, 2.1.7]). Therefore,

$$\mathbb{P}\big(\tau_n(X/\sqrt{n})^p \simeq x\big) \gtrsim \exp\big(-bn^{\frac{1}{2}+\frac{1}{p}}\delta\big).$$

The same argument can also be carried out to get the second part of the lower bound, using the deformation

$$\begin{pmatrix} 0 & \theta \\ \theta & 0 \end{pmatrix},$$

with $\theta = \left(\frac{\delta n}{2}\right)^{1/p}$.

4.5.2 Outline of proof

As for the largest eigenvalue, we use a truncation argument to isolate the entries of $X/\sqrt{n}$ which will bear the deviations of the trace $\tau_n(X/\sqrt{n})^p$, which are by the heuristics 4.5.1, the entries of order $n^{1/p}$. We decompose $X/\sqrt{n}$ in the following way

$$X/\sqrt{n} = A + B^\varepsilon + C^\varepsilon + D^\varepsilon, \tag{4.17}$$

with

$$A_{i,j} = \frac{X_{i,j}}{\sqrt{n}}\mathbb{1}_{|X_{i,j}|\leq (\log n)^d}, \qquad B^\varepsilon_{i,j} = \frac{X_{i,j}}{\sqrt{n}}\mathbb{1}_{(\log n)^d<|X_{i,j}|<\varepsilon n^{\frac{1}{2}+\frac{1}{p}}},$$

$$C^\varepsilon_{i,j} = \frac{X_{i,j}}{\sqrt{n}}\mathbb{1}_{\varepsilon n^{\frac{1}{2}+\frac{1}{p}}\leq |X_{i,j}|\leq \varepsilon^{-1}n^{\frac{1}{2}+\frac{1}{p}}}, \qquad D^\varepsilon_{i,j} = \frac{X_{i,j}}{\sqrt{n}}\mathbb{1}_{\varepsilon^{-1}n^{\frac{1}{2}+\frac{1}{p}}<|X_{i,j}|},$$

where where d is taken such that $\alpha d > 1$.

In a first phase, we will show that one can neglect in the deviations of $\tau_n(X/\sqrt{n})^p$ the contributions of the intermediate entries, that is B^ε, and the large entries, that is D^ε, so that $(\tau_n(A+C^\varepsilon)^p)_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations for $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$.

Then, due to concentration inequalities, we show that the conditional expectation given C^ε, $\mathbb{E}_{C^\varepsilon}\tau_n(H+C^\varepsilon)^p$, where H is a copy of A independent of X, are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$. From the choice of the decomposition (4.17), we deduce that C^ε has only a finite number of non-zero entries at the exponential scale $n^{1+\alpha/p}$. Thus, Lemma 4.5.1 and Wigner's theorem allow us to conclude that $(\mathbb{E}_{C^\varepsilon}\tau_n(H+C^\varepsilon)^p)_{n\in\mathbb{N}}$ is exponentially equivalent to $(\langle\mu_{sc}, x^p\rangle + \tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$. It only remains to show a large deviations principle C^ε, and conclude by contraction principle, with an argument similar as in chapter 3. The use of the contraction principle is made possible by the fact that C^ε has a finite number of non-zero entries with exponentially large probability.

4.5.3 Concentration inequalities

In this section, we revisit a concentration inequality from [79] for the trace of powers of sum of a Hermitian matrix with bounded entries with a deterministic Hermitian matrix. This inequality will be crucial to get the exponential tightness and an exponential approximation of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$.

Unfortunately, we cannot directly use the concentration inequality of [79, Proposition 4], because of the assumption made on the expectation of the entries. To make the strategy sketched in 4.5.2 work, we need to prove a concentration inequality for

$$\operatorname{tr}\left(\frac{H}{\sqrt{n}} + \frac{C}{\sqrt{n}}\right)^p,$$

where H is a centered matrix with bounded entries, and where C is a deterministic matrix whose entries are of order $n^{\frac{1}{p}+\frac{1}{2}}$. But then,

$$\operatorname{tr}\left(\frac{C}{\sqrt{n}}\right)^{2(p-1)} \leq r^{2(p-1)}n^{\frac{2(p-1)}{p}}, \tag{4.18}$$

where r is the number of non-zero entries of C, which is a bound too loose to use the concentration inequality of [79, Proposition 4].

However, since we are considering normalized traces, we are looking at deviations of order n of the traces, whereas in [79] the deviations considered were of order 1. Thus, one can expect that there is some room left in the approach of Meckes and Szarek, to get a concentration inequality for $\tau_n(H+C)^p$, with the bound (4.18).

4.5.2 Proposition. *Let $p \in \mathbb{N}$, $p \geq 3$. Let H be a centered random Hermitian matrix such that $(H_{i,j})_{i\leq j}$ are independent and bounded by some $\kappa \geq 1$, and let C be a deterministic Hermitian matrix such that $\mathrm{tr}(\frac{C}{\sqrt{n}})^{2(p-1)} \leq mn^{2-\frac{2}{p}}$, where $m \geq 1$. There are some universal constants $c, c' > 0$, such that for all $t \geq c'(pm^{p-1})^p n^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}$,*

$$\mathbb{P}\left(|\tau_n(H+C)^p - \mathbb{E}\tau_n(H+C)^p| > tn^{p/2}\right) \leq 8\exp\Big(-\frac{n^{1+\frac{2}{p}}}{c\kappa^2}h(t)\Big),$$

and

$$\mathbb{P}\left(\left|\tau_n\left|H+C\right|^p - \mathbb{E}\tau_n\left|H+C\right|^p\right| > tn^{p/2}\right) \leq 8\exp\Big(-\frac{n^{1+\frac{2}{p}}}{c\kappa^2}h(t)\Big).$$

with

$$h(t) = \min\Big\{\left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2m^{2(d-1)}}\Big\}.$$

Proof. We follow the same approach as in [79, Proposition 4], with some slight variations at times, but considering deviations of order $n^{1+p/2}$ of the trace of $(H+C)^p$. We will prove only the first inequality, the proof of the second inequality being exactly the same.

Without loss of generality, we can assume $\kappa = 1$. Let $X = H + C$. For $\beta \in \{1,2\}$, we denote by $\mathcal{H}_n^{(\beta)}$ the set of symmetric matrices of size n, when $\beta = 1$, and Hermitian matrices when $\beta = 2$. Note that as H has entries bounded by 1, we know by [70, Corollary 4.10], that for any convex and 1-Lipschitz function $f : \mathcal{H}_n^{(\beta)} \to \mathbb{R}$ with respect to the Hilbert-Schmidt norm, and all $t > 0$,

$$\mathbb{P}\left(|f(X) - \mathbb{M}f(X)| > t\right) \leq 4e^{-\frac{t^2}{4}},$$

where $\mathbb{M}f(X)$ denotes the median of $f(X)$. Let $a > 0$. Define

$$K_a = \left\{Y \in \mathcal{H}_n^{(\beta)} : ||Y||_{2(p-1)} \leq a\right\},$$

where $||Y||_q = (\mathrm{tr}|Y|^q)^{1/q}$ for any matrix Y and $q > 0$. Note that we can write

$$F = F^+ - F^-,$$

with $F^+(Y) = \mathrm{tr}Y_+^p$, and $F^-(Y) = \mathrm{tr}Y_-^p$ for any $Y \in \mathcal{H}_n^{(\beta)}$, where for every $x \in \mathbb{R}$, x_+ and x_- denote the positive and negative parts of x. The functions F^+ and F^- are convex and pa^{p-1}-Lipschitz on K_a. Let F_a^+, F_a^- denote the convex extensions of $F^+_{|K_a}$ and $F^-_{|K_a}$ to $\mathcal{H}_n^{(\beta)}$, which are pa^{p-1}-Lipschitz, as explained in [79, Lemma 5]. Then, for all $t>0$, we have

$$\mathbb{P}\left(|F_a^\sigma(X) - \mathbb{M}F_a^\sigma(X)| > tn^{1+p/2}\right) \leq 4\exp\left(-\frac{t^2 n^{p+2}}{4p^2 a^{2(d-1)}}\right),$$

with $\sigma \in \{+,-\}$.

Besides $Y \mapsto ||Y||_{2(d-1)}$ is convex and 1-Lipschitz with respect to the Hilbert-Schmidt norm. From [76, Theorem 8.6], we deduce that for any $t>0$,

$$\mathbb{P}\left(||X||_{2(p-1)} - \mathbb{E}||X||_{2(p-1)} > t\right) \leq e^{-\frac{t^2}{32}}.$$

But,

$$\mathbb{E}||X||_{2(p-1)} \leq \mathbb{E}||H||_{2(p-1)} + ||C||_{2(p-1)} \leq n^{\frac{1}{2(p-1)}}\mathbb{E}||H|| + mn^{\frac{1}{2}+\frac{1}{p}},$$

where $|| \ ||$ denotes the operator norm, and where we used the fact that $m \geq 1$. But we know that there is some universal constant $c_1 \geq 1$, such that

$$\mathbb{E}||H|| \leq c_1\sqrt{n}.$$

Thus, $\mathbb{E}||X||_{2(p-1)} \leq 2mc_1 n^{\frac{1}{2}+\frac{1}{p}}$.

Let now $b>0$, and $a = bn^{\frac{1}{2}+\frac{1}{p}}$. We have, for $b \geq 4mc_1$,

$$\mathbb{P}\Big(||X||_{2(p-1)} \geq a\Big) \leq \mathbb{P}\Big(||X||_{2(p-1)} - \mathbb{E}||X||_{2(p-1)} \geq \frac{a}{2}\Big) \leq \exp\left(-\frac{b^2 n^{1+\frac{2}{p}}}{128}\right).$$

Besides, with this choice of a, we have for all $t>0$, and all $\sigma \in \{+,-\}$,

$$\mathbb{P}\Big(|F_a^\sigma(X) - \mathbb{M}F_a^\sigma(X)| > \frac{t}{2}n^{1+p/2}\Big) \leq 4\exp\left(-\frac{t^2 n^{1+2/p}}{16p^2 b^{2(p-1)}}\right).$$

Thus,

$$\begin{aligned}\mathbb{P}\Big(|F^\sigma(X) - \mathbb{M}F_a^\sigma(X)| > \frac{t}{2}n^{1+p/2}\Big) &\leq \mathbb{P}\Big(|F_a^\sigma(X) - \mathbb{M}F_a^\sigma(X)| > \frac{t}{2}n^{1+p/2}\Big) \\ &\quad + \mathbb{P}\Big(||X||_{2(p-1)} \geq a\Big) \\ &\leq 4\exp\left(-\frac{n^{1+2/p}}{128}\min\left\{b^2, \frac{t^2}{p^2 b^{2(p-1)}}\right\}\right).\end{aligned}$$

As a consequence, for $b = 4mc_1$, we can find a numerical constant $c_2 \geq 1$, such that for $t = c_2 pn^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}$, we have

$$\mathbb{P}\Big(F^\sigma(X) - \mathbb{M}F_a^\sigma(X) > tn^{1+p/2}\Big) < \frac{1}{2}.$$

We deduce that

$$\mathbb{M}F^\sigma(X) \leq \mathbb{M}F_a^\sigma(X) + c_2 pn^{\frac{1}{2}+\frac{p}{2}-\frac{1}{p}}.$$

As F_a^σ is non-decreasing with a, and $F_a^\sigma \leq F^\sigma$ for any $a > 0$, we have for all $b \geq 4mc_1$,

$$\mathbb{M}F^\sigma(X) - c_2 p n^{\frac{1}{2}+\frac{p}{2}-\frac{1}{p}} \leq \mathbb{M}F_a^\sigma(X) \leq \mathbb{M}F^\sigma(X).$$

Thus, for $t \geq 2c_2 p n^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}$, and any $b \geq 4mc_1$, we deduce that

$$\begin{aligned}\mathbb{P}\left(|F^\sigma(X) - \mathbb{M}F^\sigma(X)| > tn^{1+p/2}\right) &\leq \mathbb{P}\Big(|F^\sigma(X) - \mathbb{M}F_a^\sigma(X)| > \frac{t}{2}n^{1+p/2}\Big)\\ &\leq 4\exp\left(-\frac{n^{1+2/p}}{128}\min\left\{b^2, \frac{t^2}{p^2 b^{2(p-1)}}\right\}\right).\end{aligned}$$

But one can check that,

$$\max_{b\geq 4mc_1} \min\left\{b^2, \frac{t^2}{p^2 b^{2(p-1)}}\right\} = \min\left\{\left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2(mc_1)^{2(p-1)}}\right\}.$$

Optimizing in b in the previous inequality, and setting $c_3 = 128c_1^{2(p-1)}$, we get

$$\mathbb{P}\Big(|F^\sigma(X) - \mathbb{M}F^\sigma(X)| > tn^{1+p/2}\Big) \leq 4\exp\left(-\frac{n^{1+\frac{2}{p}}}{c_3}\min\Big\{\left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2 m^{2(p-1)}}\Big\}\right).$$

To get the same inequality but with $\mathbb{E}F^\sigma(X)$ instead of $\mathbb{M}F^\sigma(X)$, we integrate by parts the inequality above, and we find that there is some constant $c_4 > 0$, such that

$$|\mathbb{E}F^\sigma(X) - \mathbb{M}F^\sigma(X)| \leq c_4 m^{p-1} p n^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}.$$

At the price of taking c_4 larger, we can assume that $c_4 \geq c_2$. Then, for every $t \geq 2c_4 m^{p-1} p n^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}$,

$$\begin{aligned}\mathbb{P}\Big(|F^\sigma(X) - \mathbb{E}F^\sigma(X)| > tn^{1+p/2}\Big) &\leq \mathbb{P}\Big(|F^\sigma(X) - \mathbb{M}F^\sigma(X)| > \frac{t}{2}n^{1+p/2}\Big)\\ &\leq 4\exp\left(-\frac{n^{1+\frac{2}{p}}}{4c_3}\min\Big\{\left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2 m^{2(d-1)}}\Big\}\right).\end{aligned}$$

As $F = F^+ - F^-$, we have for any $t \geq 2c_4 m^{p-1} p n^{-\frac{1}{2}\left(1+\frac{2}{p}\right)}$,

$$\mathbb{P}\left(|F(X) - \mathbb{E}F(X)| > tn^{1+p/2}\right) \leq 8\exp\left(-\frac{n^{1+\frac{2}{p}}}{16c_3}\min\left\{\left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{m^{2(d-1)}}\right\}\right).$$

Setting $c = 16c_3$, and $c' = 2c_4$, we get the claim. □

4.5.4 Exponential tightness

Throughout the rest of this section, we fix a constant $\gamma > 0$, such that for t large enough,

$$\mathbb{P}\left(|X_{1,1}| > t\right) \vee \mathbb{P}\left(|X_{1,2}| > t\right) \leq e^{-\gamma t^\alpha}. \tag{4.19}$$

In this section, we will show that the sequence $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$ is exponentially tight, namely, we have the following proposition.

4.5.3 Proposition (Exponential tightness)**.**

$$\lim_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log \mathbb{P}\left(\tau_n \left|X/\sqrt{n}\right|^p > t\right) = -\infty.$$

Proof of Proposition 4.5.3. Using the triangular inequality for the p-Schatten norm, we get for any $t>0$,

$$\begin{aligned}\mathbb{P}\left(\tau_n \left|X/\sqrt{n}\right|^p > (4t)^p\right) &\leq \mathbb{P}\left(\tau_n |A|^p > t^p\right) + \mathbb{P}\left(\tau_n |B^\varepsilon|^p > t^p\right)\\ &\quad + \mathbb{P}\left(\tau_n |C^\varepsilon|^p > t^p\right) + \mathbb{P}\left(\tau_n |D^\varepsilon|^p > t^p\right). \end{aligned} \tag{4.20}$$

This shows that it suffices to estimate at the exponential scale, the probability of each event $\{\tau_n |A|^p > t^p\}$, $\{\tau_n |B^\varepsilon|^p > t^p\}$, $\{\tau_n |C^\varepsilon|^p > t^p\}$, and finally $\{\tau_n |D^\varepsilon|^p > t^p\}$. As a consequence of the concentration inequality of Proposition 4.5.2, we have the following lemma.

4.5.4 Lemma.

$$\lim_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log \mathbb{P}\left(\tau_n |A|^p > t\right) = -\infty,$$

where A is as in (4.17).

Proof. Note that as $p\geq 2$,

$$\operatorname{tr}(\mathbb{E}A)^{2(p-1)} \leq \left(\operatorname{tr}(\mathbb{E}A)^2\right)^{p-1}.$$

Since the entries of X are centered, we get

$$\operatorname{tr}(\mathbb{E}A)^2 = \frac{1}{n}\sum_{1\leq i,j\leq n} \mathbb{E}|X_{i,j}|^2 \mathbb{1}_{|X_{i,j}|>(\log n)^d}.$$

Integrating by parts, we have

$$\operatorname{tr}(\mathbb{E}A)^2 = O\Big(n^2 e^{-\frac{\gamma}{2}(\log n)^{\alpha d}}\Big),$$

where γ is as in (4.19). As $\alpha d > 1$,

$$\operatorname{tr}(\mathbb{E}A)^{2(p-1)} = o(1). \tag{4.21}$$

We see that A satisfies the assumptions of Proposition 4.5.2 with some $m\geq 1$ and $\kappa = (\log n)^d$. We get for any $t>0$, and N large enough,

$$\mathbb{P}\left(|\tau_n |A|^p - \mathbb{E}\tau_n |A|^p| > t\right) \leq 8\exp\left(-\frac{n^{1+\frac{2}{p}}}{cp^2(\log n)^{2d}}\min\left\{t^{2/p}, \frac{t^2}{m^{2(p-1)}}\right\}\right),$$

which yields, as $\alpha<2$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log \mathbb{P}\left(|\tau_n |A|^p - \mathbb{E}\tau_n |A|^p| > t\right) = -\infty. \tag{4.22}$$

We know from [3, Theorem 2.1.1, Lemma 2.1.6], that

$$\mathbb{E}\tau_n |X/\sqrt{n}|^p \underset{n\to+\infty}{\longrightarrow} \langle \mu_{sc}, |x|^p\rangle, \tag{4.23}$$

where $\langle \mu_{sc}, |x|^p\rangle = \int |x|^p d\mu_{sc}(x)$. Denoting $\mu_{X/\sqrt{n}}$ and μ_A the spectral measures of $X/\sqrt{n}$ and A respectively, we have using the decreasing coupling and [22, Theorem III 4.4],

$$\mathcal{W}_p(\mathbb{E}\mu_{X/\sqrt{n}}, \mathbb{E}\mu_A) \leq \Big(\mathbb{E}\tau_n|X/\sqrt{n} - A|^p\Big)^{1/p}, \tag{4.24}$$

where $\mathcal{W}_p$ is the L^p-Wasserstein distance. As a consequence of the polar decomposition, we can write $|X/\sqrt{n} - A|^p = (X/\sqrt{n} - A)^p U$, where U is a unitary matrix, so that

$$\mathbb{E}\mathrm{tr}|X/\sqrt{n} - A|^p \leq \frac{1}{n^{p/2}} \sum_{i_1,\dots,i_{p+1}} \mathbb{E}\prod_{j=1}^{p} |X_{i_j,i_{j+1}}|\mathbb{1}_{|X_{i_j,i_{j+1}}|\leq (\log n)^d}, \tag{4.25}$$

Hölder inequality yields,

$$\mathbb{E}\mathrm{tr}|X/\sqrt{n} - A|^p \leq n^{p/2+1}\max\Big(\mathbb{E}|X_{1,1}|^p\mathbb{1}_{|X_{1,1}|>(\log n)^p}, \mathbb{E}|X_{1,2}|^p\mathbb{1}_{|X_{1,2}|>(\log n)^p}\Big),$$

where we used the fact that the entries of X are centered. Integrating by parts, we get

$$\mathbb{E}\mathrm{tr}|X/\sqrt{n} - A|^p = O\Big(n^{p/2+1}e^{-\frac{\gamma}{2}(\log n)^{\alpha d}}\Big), \tag{4.26}$$

where γ is as in (4.19). As $\alpha d > 1$, we deduce by (4.24), $\mathcal{W}_p(\mathbb{E}\mu_{X/\sqrt{n}}, \mathbb{E}\mu_A) = o(1)$, which yields

$$\Big|\mathbb{E}\tau_n|X/\sqrt{n}|^p - \mathbb{E}\tau_n|A|^p\Big| = o(1).$$

We can conclude with (4.22) and (4.23) that $(\tau_n|A|^p)_{n\in\mathbb{N}}$ is exponentially tight. □

For the second event $\{\tau_n|B^\varepsilon|^p > t^p\}$, we have the following lemma.

4.5.5 Lemma. *For any $\varepsilon > 0$, we have*

$$\lim_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\tau_n|B^\varepsilon|^p > t\right) = -\infty.$$

Proof. Since $p \geq 2$, we have

$$(\mathrm{tr}\,|B^\varepsilon|^p)^{2/p} \leq \mathrm{tr}(B^\varepsilon)^2.$$

Thus,

$$\mathbb{P}\left(\mathrm{tr}|B^\varepsilon|^p \geq tn\right) \leq \mathbb{P}\left(\mathrm{tr}(B^\varepsilon)^2 \geq t^{2/p}n^{2/p}\right).$$

Chernoff's inequality yields for any $\lambda > 0$,

$$\mathbb{P}\Big(\sum_{1\leq i\leq j\leq n}\big|B^\varepsilon_{i,j}\big|^2 \geq \frac{t^{2/p}}{2}n^{2/p}\Big) \leq e^{-\frac{\lambda}{2}t^{\frac{2}{p}}n^{\frac{2}{p}+1}}\prod_{1\leq i\leq j\leq n}\mathbb{E}\left(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log n)^d<|X_{i,j}|<\varepsilon n^{\frac{1}{2}+\frac{1}{p}}}}\right).$$

Let $1 \leq i \leq j \leq n$. By the integration by part formula (3.17), we get for n large enough,

$$\mathbb{E}\left(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log n)^d<|X_{i,j}|<\varepsilon n^{\frac{1}{2}+\frac{1}{p}}}}\right) \leq 1 + \int_{(\log n)^d}^{\varepsilon n^{\frac{1}{2}+\frac{1}{p}}} 2\lambda x e^{f(x)}dx,$$

with $f(x) = \lambda x^2 - \gamma x^\alpha$, and γ is as in (4.19). Let

$$\lambda = \frac{\alpha\gamma}{2}\varepsilon^{\alpha-2}n^{-(2-\alpha)\left(\frac{1}{2}+\frac{1}{p}\right)}.$$

With this choice of λ, one can easily check that f is non-increasing on $[(\log n)^d, \varepsilon n^{\frac{1}{2}+\frac{1}{p}}]$. Thus,

$$\begin{aligned}\mathbb{E}\left(e^{\lambda|X_{i,j}|^2 \mathbb{1}_{(\log n)^d < |X_{i,j}| < \varepsilon n^{\frac{1}{2}+\frac{1}{p}}}}\right) &\leq 1 + 2\lambda\varepsilon^2 n^{1+\frac{2}{p}} e^{f\left((\log n)^d\right)} \\ &\leq 1 + \alpha\gamma\varepsilon^\alpha n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)} e^{f\left((\log n)^d\right)}.\end{aligned}$$

But for n large enough,

$$f((\log n)^d) = \frac{\alpha\gamma}{2}\varepsilon^{\alpha-2}n^{-(2-\alpha)\left(\frac{1}{2}+\frac{1}{p}\right)}(\log n)^{2d} - \gamma(\log n)^{\alpha d} \leq -\frac{\gamma}{2}(\log n)^{\alpha d}.$$

As $\alpha d > 1$, we get for N large enough,

$$\mathbb{E}\left(e^{\lambda|X_{i,j}|^2 \mathbb{1}_{(\log n)^d < |X_{i,j}| < \varepsilon n^{\frac{1}{2}+\frac{1}{p}}}}\right) \leq 1 + e^{-\frac{\gamma}{4}(\log n)^{\alpha d}} \leq \exp\left(e^{-\frac{\gamma}{4}(\log n)^{\alpha d}}\right).$$

Then,

$$\mathbb{P}\left(\operatorname{tr}|B^\varepsilon|^p \geq tn\right) \leq \exp\left(-\frac{\alpha\gamma}{4}\varepsilon^{\alpha-2}n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}t^{\frac{2}{p}}\right)\exp\left(n^2 e^{-\frac{\gamma}{2}(\log n)^{\alpha d}}\right). \tag{4.27}$$

Since $\alpha d > 1$, we get

$$\lim_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\operatorname{tr}|B^\varepsilon|^p \geq tn\right) = -\infty.$$

□

We now turn to the event $\{\tau_n|C^\varepsilon|^p > t\}$. As a consequence of Bennett's inequality, we have the following lemma.

4.5.6 Lemma. *For any $\varepsilon > 0$,*

$$\lim_{t\to+\infty}\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\tau_n|C^\varepsilon|^p > t\right) = -\infty.$$

To prove this lemma, we will first show that at the exponential scale C^ε has a finite number of non-zero entries.

4.5.7 Proposition. *For all $\varepsilon > 0$,*

$$\lim_{r\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\operatorname{Card}\left\{(i,j) : C^\varepsilon_{i,j} \neq 0\right\} \geq r\right) = -\infty,$$

where C^ε is as in (4.17).

Proof. Let $\varepsilon > 0$. Note that

$$\mathbb{P}\left(\mathrm{Card}\left\{(i,j) : C^{\varepsilon}_{i,j} \neq 0\right\} \geq r\right) \leq \mathbb{P}\left(\sum_{1\leq i\leq j\leq n} \mathbb{1}_{|X_{i,j}|\geq \varepsilon n^{\frac{1}{2}+\frac{1}{p}}} \geq \frac{r}{2}\right).$$

Let $p_{i,j} = \mathbb{P}\left(|X_{i,j}| \geq \varepsilon n^{\frac{1}{2}+\frac{1}{p}}\right)$, for $i,j \in \{1,2\}$. From (4.19), we have

$$p_{1,1} \vee p_{1,2} = o\left(\frac{1}{n^2}\right).$$

Therefore, it is enough to show that

$$\lim_{r\to+\infty} \limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\Big(\sum_{1\leq i\leq j\leq n} (\mathbb{1}_{|X_{i,j}|\geq \varepsilon n^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \geq r\Big) = -\infty.$$

By Bennett's inequality (see [76, Theorem 2.9]) we have,

$$\mathbb{P}\left(\sum_{1\leq i\leq j\leq n} (\mathbb{1}_{|X_{i,j}|\geq \varepsilon n^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \geq r\right) \leq \exp\Big(-vh\Big(\frac{r}{v}\Big)\Big),$$

with $h(x) = (x+1)\log(x+1) - x$, and $v = \sum_{i\leq j} p_{i,j}$. From (4.19), we have for N large enough,

$$v \leq n^2 e^{-\gamma\varepsilon^{\alpha} n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}}.$$

As $h(x) \underset{+\infty}{\sim} x\log(x)$, we get for n large enough,

$$\mathbb{P}\left(\sum_{1\leq i,j\leq n} (\mathbb{1}_{|X_{i,j}|\geq \varepsilon n^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \geq r\right) \leq \exp\big(-r\gamma\varepsilon^{\alpha} n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}\big) \exp\Big(r\log\Big(\frac{r}{n^2}\Big)\Big),$$

which gives the claim. $\square$

With this result on the number of non-zero entries of C^{ε}, we will see that the matrix $\frac{1}{n}|C^{\varepsilon}|^p$ has a finite number of non-zero entries of order 1, and that it yields the exponential estimate claimed in Lemma 4.5.6.

Proof of Lemma 4.5.6. Using the polar decomposition as in (4.25), and bounding each coefficient of C^{ε} by $\varepsilon^{-1}n^{1/p}$, we get,

$$\mathrm{tr}\,|C^{\varepsilon}|^p \leq |\mathcal{I}^{\varepsilon}|^p n\varepsilon^{-p},$$

where $|\mathcal{I}^{\varepsilon}|$ denotes the number of non-zero entries in C^{ε}. Due to Lemma 4.5.7, we get,

$$\lim_{t\to+\infty} \limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\tau_n\,|C^{\varepsilon}|^p > t\right) = -\infty.$$

$\square$

At last, we prove the following exponential tightness for $\tau_n|D^{\varepsilon}|^p$.

4.5.8 Lemma. *It holds*

$$\lim_{\varepsilon\to 0}\limsup_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\tau_n\left|D^\varepsilon\right|^p > t\right) = -\infty,$$

with D^ε as in (4.17).

Proof. A union bound gives for n large enough,

$$\mathbb{P}\left(D^\varepsilon \neq 0\right) \leq n^2 \exp\Big(-\gamma\varepsilon^{-\alpha} n^{\alpha(\frac{1}{2}+\frac{1}{p})}\Big), \tag{4.28}$$

with γ as in (4.19). □

From (4.20), lemmas 4.5.4, 4.5.5, and 4.5.6, we get for any $\varepsilon > 0$,

$$\begin{aligned}\limsup_{t\to+\infty}\limsup_{n\to+\infty} & n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\tau_n\left|X/\sqrt{n}\right|^p > t\right)\\ &\leq \limsup_{t\to+\infty}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\tau_n|D^\varepsilon|^p > t\right).\end{aligned}$$

Taking the limsup as ε goes to 0, we see that Lemma 4.5.8 yields the exponential tightness claimed in Proposition 4.5.3. □

4.5.5 Exponential equivalences

4.5.6 First step

We will prove in this section that we can ignore in the deviations of $\tau_n(X/\sqrt{n})^p$ the contributions of the large entries, namely those such that $|X_{i,j}| > \varepsilon^{-1}n^{\frac{1}{2}+\frac{1}{p}}$, and the contributions of the intermediate entries, that is $(\log n)^d < |X_{i,j}| < \varepsilon n^{\frac{1}{2}+\frac{1}{p}}$. More precisely, we will prove the following exponential approximation.

4.5.9 Proposition. *For any $t > 0$,*

$$\lim_{\varepsilon\to 0}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\left|\tau_n(X/\sqrt{n})^p - \tau_n\left(A + C^\varepsilon\right)^p\right| > t\right) = -\infty,$$

with A and C^ε are as in (4.17)*. In other words, $(\tau_n(A+C^\varepsilon)^p)_{n\in\mathbb{N}}$ are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$.*

Proof. Let $\tau > 0$. Define the compact subset,

$$\mathcal{K}_\tau = \{\mu \in \mathcal{P}(\mathbb{R}) : \langle \mu, |x|^p\rangle \leq \tau\}.$$

As the function which associates to a probability measure μ on $\mathbb{R}$, its p^{th} moment, $\langle \mu, x^p\rangle$, is continuous for the L^p-Wasserstein distance, we get that restricted to $\mathcal{K}_\tau$, it is uniformly continuous. Applying this uniform continuity to spectral measures of Hermitian matrices, and using Lemma 2.4.1, we get that there exists a non-negative function h depending on τ, satisfying $h(t) \to 0$ as $t \to 0$, such that for any $X, Y \in \mathcal{H}_n^{(\beta)}$, if

$$\tau_n|X|^p \leq \tau, \text{ and } |\tau_n X^p - \tau_n Y^p| > t,$$

for some $t > 0$, then,

$$\tau_n\, |X - Y|^p > h(t).$$

But, from Proposition 4.5.3, we know that $(\tau_n|X/\sqrt{n}|^p)_{n\in\mathbb{N}}$ is exponentially tight, therefore, it is enough to show that for any $\tau > 0$,

$$\lim_{\varepsilon\to 0}\limsup_{n\to+\infty} n^{-\alpha\left(\frac{1}{2}+\frac{1}{p}\right)} \log \mathbb{P}\left(\left|\tau_n(X/\sqrt{n})^p - \tau_n(A+C^\varepsilon)^p\right| > t, \tau_n|X/\sqrt{n}|^p \leq \tau\right) = -\infty.$$

Let $\tau > 0$. With the previous observation, we get for any $t > 0$,

$$\mathbb{P}\left(\left|\tau_n(X/\sqrt{n})^p - \tau_n(A+C^\varepsilon)^p\right| > t, \tau_n|X/\sqrt{n}|^p \leq \tau\right) \leq \mathbb{P}\Big(\tau_n|B^\varepsilon + D^\varepsilon|^p > h(t)\Big).$$

By the triangular inequality for the p-Schatten norm, we get

$$\begin{aligned}\mathbb{P}\Big(|\tau_n(X/\sqrt{n})^p - &\tau_n(A+C^\varepsilon)^p| > t, \tau_n|X/\sqrt{n}|^p \leq \tau\Big)\\ &\leq \mathbb{P}\left(\tau_n|B^\varepsilon|^p > \frac{h(t)}{2^p}\right) + \mathbb{P}\left(\tau_n|D^\varepsilon|^p > \frac{h(t)}{2^p}\right). \qquad (4.29)\end{aligned}$$

But, on one hand (4.27) yields

$$\lim_{\varepsilon\to 0}\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\tau_n|B^\varepsilon|^p > \frac{h(t)}{2^p}\right) = -\infty,$$

and on the other hand, (4.28) gives

$$\lim_{\varepsilon\to 0}\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\tau_n|D^\varepsilon|^p > \frac{h(t)}{2^p}\right) = -\infty.$$

This concludes the proof of Proposition 4.5.9, taking the limsup as n goes to $+\infty$ at the exponential scale, and then the limsup as ε goes to 0 in (4.29).

□

4.5.7 Second step

We show here that in the study of the deviations of $\tau_n(A+C^\varepsilon)^p$, we can replace A by a matrix H independent of X, and that $\tau_n(H+C^\varepsilon)^p$ is exponentially equivalent to its conditional expectation given the σ-algebra $\mathcal{F}$, generated by the $X_{i,j}$ such that $|X_{i,j}| > (\log n)^d$. More precisely, we will prove the following result.

4.5.10 Proposition. *Let $\mathcal{F}$ be the σ-algebra generated by the variables $X_{i,j}\mathbb{1}_{|X_{i,j}|>(\log n)^d}$. Let H be a random Hermitian matrix independent of X, such that $(H_{i,j})_{i\leq j}$ are independent, and for all $1 \leq i \leq n$, $H_{i,i}$ has the same law as $X_{1,1}/\sqrt{n}$ conditioned on $\{|X_{1,1}| \leq (\log n)^d\}$, and for all $i < j$, $H_{i,j}$ has the same law as $X_{1,2}/\sqrt{n}$ conditioned on $\{|X_{1,2}| \leq (\log n)^d\}$.*

For any $t > 0$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left|\mathrm{tr}(X/\sqrt{n})^p - \mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(H+C^\varepsilon)^p\right| > tn\right) = -\infty,$$

where $\mathbb{E}_{\mathcal{F}}$ denotes the conditional expectation given $\mathcal{F}$.

Proof. By Proposition 4.5.9, we know that $(\tau_n(A+C^\varepsilon)^p)_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$, therefore it is enough to show that for all $\varepsilon>0$, and $t>0$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(|\mathrm{tr}\,(A+C^\varepsilon)^p-\mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(H+C^\varepsilon)^p|>tn\right)=-\infty.$$

From Proposition 4.5.6, we see that is actually sufficient to show that for any $r\in\mathbb{N}$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(|\mathrm{tr}\,(A+C^\varepsilon)^p-\mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(H+C^\varepsilon)^p|>tn,|\mathcal{I}_\varepsilon|\le r\right)=-\infty,$$

where

$$\mathcal{I}_\varepsilon=\Big\{(i,j)\in\{1,\ldots,n\}\times\{1,\ldots,n\}:C^\varepsilon_{i,j}\neq 0\Big\}.$$

Note that C^ε is $\mathcal{F}$-measurable, and given $\mathcal{F}$, A has independent up-diagonal entries bounded by $(\log n)^d/\sqrt{n}$. Moreover, using the triangle inequality for the $2(p-1)$-Schatten norm, and the convexity, we get

$$\mathrm{tr}(\mathbb{E}A+C^\varepsilon)^{2(p-1)}\le 2^{2(p-1)}\max\Big(\mathrm{tr}(\mathbb{E}A)^{2(p-1)},\mathrm{tr}(C^\varepsilon)^{2(p-1)}\Big).$$

On one hand, we have, expanding the trace and bounding each entry of C^ε by $\varepsilon^{-1}n^{1/p}$,

$$\mathrm{tr}(C^\varepsilon)^{2(p-1)}\le|\mathcal{I}_\varepsilon|^{2(p-1)}\varepsilon^{-2(p-1)}n^{2-\frac{2}{p}},$$

and on the other hand we have from (4.21) that $\mathrm{tr}(\mathbb{E}A)^{2(p-1)}=o(1)$. Therefore, we can apply the result of Proposition 4.5.2 for the trace of $(A+C^\varepsilon)^p$ under the conditional probability given $\mathcal{F}$. As $\alpha<2$, we get that for any $t>0$, and $r\in\mathbb{N}$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(|\mathrm{tr}\,(A+C^\varepsilon)^p-\mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(A+C^\varepsilon)^p|>tn,|\mathcal{I}_\varepsilon|\le r\right)=-\infty.$$

We will use the same decoupling argument as in [29], to remove the dependency between A and C^ε. Let $I=\big\{(i,j):|X_{i,j}|>(\log n)^d\big\}$. Define A' the $n\times n$ matrix with (i,j)-entry

$$A'_{i,j}=A_{i,j}\mathbb{1}_{(i,j)\notin I}+H_{i,j}\mathbb{1}_{(i,j)\in I}. \tag{4.30}$$

Note that A' and H are both independent of $\mathcal{F}$ and have the same law. Therefore,

$$\mathbb{E}_{\mathcal{F}}\mathrm{tr}\left(A'+C^\varepsilon\right)^p=\mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(H+C^\varepsilon)^p.$$

Due to the triangular inequality and Lemma 4.5.6, it only remains to prove that for any $t>0$, and any $\tau>0$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(\left|\mathbb{E}_{\mathcal{F}}\mathrm{tr}\,(A+C^\varepsilon)^p-\mathbb{E}_{\mathcal{F}}\mathrm{tr}\left(A'+C^\varepsilon\right)^p\right|>tn,\tau_n|C^\varepsilon|^p\le\tau\right)=-\infty.$$

But, using again the triangular inequality for the p-Schatten norm, we get

$$\mathbb{E}_{\mathcal{F}}\tau_n|A'+C^\varepsilon|^p\le 2^p\max\left(\mathbb{E}\tau_n|H|^p,\tau_n|C^\varepsilon|^p\right).$$

With the same argument as in the proof of Lemma 4.5.4 we have

$$\mathbb{E}\tau_n|H|^p\underset{n\to+\infty}{\longrightarrow}\langle\mu_{sc},|x|^p\rangle.$$

Arguing as in the proof of Lemma 4.5.4, we see that it is sufficient to show that for any $t > 0$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\mathcal{W}_p(\mathbb{E}_{\mathcal{F}}\mu_{A+C^\varepsilon}, \mathbb{E}_{\mathcal{F}}\mu_{A'+C^\varepsilon}) > t\right) = -\infty,$$

where μ_{A+C^ε} and $\mu_{A'+C^\varepsilon}$ denote the spectral measures of $A+C^\varepsilon$ and $A'+C^\varepsilon$. But,

$$\mathcal{W}_p(\mathbb{E}_{\mathcal{F}}\mu_{A+C^\varepsilon}, \mathbb{E}_{\mathcal{F}}\mu_{A'+C^\varepsilon})^p \leq \mathbb{E}_{\mathcal{F}}\tau_n|A-A'|^p,$$

and besides, expanding the trace using the polar decomposition, we get

$$\mathbb{E}_{\mathcal{F}}\tau_n \left|A-A'\right|^p \leq c_0 \frac{|I|^p}{n^{1+p/2}}, \tag{4.31}$$

where c_0 is constant independent of N such that,

$$\max\left(\mathbb{E}\left|\sqrt{n}H_{1,1}\right|^p, \mathbb{E}\left|\sqrt{n}H_{1,2}\right|^p\right) \leq c_0.$$

Thus, in order to control $\mathbb{E}_{\mathcal{F}}\tau_n\left|A-A'\right|^p$, we need to make sure that I contains no more than $tn^{1+p/2}$ indices, for any $t > 0$, at the exponential scale $n^{\alpha(\frac{1}{2}+\frac{1}{p})}$. By a argument similar as in the proof of Proposition 4.5.7, we get the following lemma.

4.5.11 Lemma. *Let $I = \left\{(i,j) : |X_{i,j}| > (\log n)^d\right\}$. For $\delta > 0$, we define the event,*

$$F_\delta = \left\{|I| \leq \frac{\delta}{c_0} n^{1+2/p}\right\}.$$

It holds that

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(F_\delta^c\right) = -\infty.$$

Using (4.31), and Lemma (4.5.11), we get the claim.

□

4.5.8 Third step

We showed in Proposition 4.5.10 that $(\mathbb{E}_{\mathcal{F}}\tau_n\left(H+C^\varepsilon\right)^p)_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$. We will prove now that we can approximate $\mathbb{E}_{\mathcal{F}}\tau_n\left(H+C^\varepsilon\right)^p$ at the exponential scale $n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$, by $\mathbb{E}\tau_n H^p + \tau_n(C^\varepsilon)^p$, and then by $\langle\mu_{sc}, x^p\rangle + \tau_n(C^\varepsilon)^p$. This will give good exponential approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$, as stated in the following proposition.

4.5.12 Proposition. *For any $t > 0$,*

$$\lim_{\varepsilon\to 0}\limsup_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left|\tau_n(X/\sqrt{n})^p - \langle\mu_{sc}, x^p\rangle - \tau_n\left(C^\varepsilon\right)^p\right| > t\right) = -\infty,$$

where A and C^ε are as in (4.17).

In order to prove that $\mathbb{E}\tau_n H^p + \tau_n(C^\varepsilon)^p$ is an exponential equivalent of $\mathbb{E}_{\mathcal{F}}\tau_n\left(H+C^\varepsilon\right)^p$, we will need the following deterministic lemma.

4.5.13 Lemma. *Let $p \geq 2$. Let H and C be two Hermitian matrices of size N. Assume that C is of rank at most r. We have*

$$|\mathrm{tr}\,(H+C)^p - \mathrm{tr}H^p - \mathrm{tr}C^p| \leq 2^p r \max_{1\leq k\leq p-1} ||H||^k ||C||^{p-k},$$

where $||\ ||$ denotes the operator norm.

Proof. Expanding the sum we get

$$\mathrm{tr}\,(H+C)^p = \sum_{k=0}^{p} \sum_{\substack{M^{(i)}\in\{H,C\} \\ |\{i:M^{(i)}=H\}|=k}} \mathrm{tr}(M^{(1)}...M^{(p)}).$$

Let $k \in \{1,...,p-1\}$, and let $M^{(1)}$,...,$M^{(p)}$ be matrices such that $M^{(i)} \in \{H, C\}$, and $\mathrm{Card}\{i : M^{(i)} = H\} = k$. Let $(\eta_j)_{1\leq j\leq N}$ be an orthonormal basis of eigenvectors for C such that $\eta_{r+1}, ..., \eta_N$ are in the kernel of C. Using the cyclicity of the trace, we can assume $M^{(p)} = C$. Assuming $M^{(p)} = C$, we get

$$\begin{aligned}\left|\mathrm{tr}\left(M^{(1)}...M^{(p)}\right)\right| &= \left|\sum_{j=1}^{n}\langle M^{(1)}...M^{(p)}\eta_j,\eta_j\rangle\right| \\ &= \left|\sum_{j=1}^{r}\langle M^{(1)}...M^{(p)}\eta_j,\eta_j\rangle\right| \\ &\leq r||H||^k||C||^{p-k},\end{aligned}$$

which ends the proof of the claim.

□

Proof. Note that the same argument as in the proof of Lemma 4.5.4 yields

$$\mathbb{E}\tau_n H^p \underset{n\to+\infty}{\longrightarrow} \langle \mu_{sc}, x^p\rangle\,,$$

Therefore, due to Proposition 4.5.10, we only need to prove that for any $\varepsilon > 0$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(|\mathbb{E}_{\mathcal{F}}\tau_n\,(H+C^\varepsilon)^p - \mathbb{E}\tau_n H^p - \tau_n\,(C^\varepsilon)^p| > t\right) = -\infty.$$

Using Lemma 4.5.1 and the fact that the rank of a matrix is bounded by the number of its non-zero entries, we have

$$|\mathbb{E}_{\mathcal{F}}\tau_n\,(H+C^\varepsilon)^p - \mathbb{E}\tau_n H^p - \tau_n\,(C^\varepsilon)^p| \leq \frac{2^p}{n}|\mathcal{I}_\varepsilon| \max_{1\leq k\leq p-1}\left\{||C^\varepsilon||^{p-k}\mathbb{E}||H||^k\right\},$$

where $\mathcal{I}_\varepsilon$ denotes the set of indices (i,j) such that $C^\varepsilon_{i,j} \neq 0$. But,

$$||C^\varepsilon|| \leq |\mathcal{I}_\varepsilon| \sup_{i,j}|C_{i,j}| \leq |\mathcal{I}_\varepsilon|\,\varepsilon^{-1}n^{1/p}.$$

Thus,

$$|\mathbb{E}_{\mathcal{F}}\tau_n\,(H+C^\varepsilon)^p - \mathbb{E}\tau_n H^p - \tau_n\,(C^\varepsilon)^p| \leq \frac{2^p\varepsilon^{-p+1}}{n^{1/p}}|\mathcal{I}_\varepsilon|^p \max_{1\leq k\leq p-1}\mathbb{E}||H||^k.$$

But we know from [3, Theorem 2.1.22, Exercice 2.1.27] that $||X/\sqrt{n}||$ converges in all L^p spaces to 2, and we have

$$\mathbb{E}||X/\sqrt{n} - H||^p = \mathbb{E}||X/\sqrt{n} - A'||^p \leq \mathbb{E}\mathrm{tr}|X/\sqrt{n} - A'|^p,$$

where A' is as in (4.30). With the same argument as in Lemma 4.5.4, we get

$$\mathbb{E}\mathrm{tr}|X/\sqrt{n} - A'|^p = o(1).$$

Thus, for any $k \in \{1, ..., p\}$, $\mathbb{E}||H||^k$ is bounded. We can find a constant $M_p > 0$ such that,

$$|\mathbb{E}_{\mathcal{F}}\tau_n (A + C^\varepsilon)^p - \mathbb{E}\tau_n A^p - \tau_n (C^\varepsilon)^p| \leq M_p|\mathcal{I}_\varepsilon|^p n^{-\frac{1}{p}}.$$

Thus, for any $t > 0$, and $r \in \mathbb{N}$,

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(|\mathbb{E}_{\mathcal{F}}\tau_n (A + C^\varepsilon)^p - \mathbb{E}\tau_n A^p - \tau_n (C^\varepsilon)^p| > t, |\mathcal{I}_\varepsilon| \leq r\right) = -\infty.$$

Invoking Lemma 4.5.7, we get the claim. □

4.5.9 A large deviations principle for $\tau_n(X/\sqrt{n})^p$

We proved in the previous section that $(\langle\mu_{sc}, x^p\rangle + \tau_n(C^\varepsilon)^p)_{\varepsilon>0,n\in\mathbb{N}}$ are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$ at the exponential scale considered. The aim of this section is to show that we can derive a LDP for each $\varepsilon > 0$ for $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$, using the contraction principle, and deduce a LDP for $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$.

In the view of applying a contraction principle for the sequence $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$, we need to find a good space to embed C^ε so that we can define a trace which will be continuous. As in the proof of the LDP of the largest eigenvalue (see chapter 3 §3.10), we define for every $r \in \mathbb{N}$,

$$\mathcal{E}_r = \{A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)} : \mathrm{Card}\{(i,j) : A_{i,j} \neq 0\} \leq r\}.$$

We denote $\widetilde{\mathcal{E}}_r$ the set of equivalence classes of $\mathcal{E}_r$ under the action of $\mathfrak{S}_r$. We equip $\mathcal{E}_r$ of the topology inherited from $\mathcal{H}_r^{(\beta)}/\mathfrak{S}_r$, as $\mathcal{E}_r$ can be seen as a subset of $\mathcal{H}_r^{(\beta)}/\mathfrak{S}_r$. This topology is metrizable by the distance $\tilde{d}$ defined in (3.48).

Since the trace is continuous and invariant by conjugation, we can define the trace on $\mathcal{H}_r^{(\beta)}/\mathfrak{S}_r$ and it will be still continuous. Therefore, the trace on $\widetilde{\mathcal{E}}_r$ is continuous for the topology we defined above.

Let $\varepsilon > 0$. Let $\mathbb{P}^\varepsilon_{n,r}$ denote the law of $C^\varepsilon/n^{1/p}$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$, and $\widetilde{\mathbb{P}}^\varepsilon_{n,r}$ the push-forward of $\mathbb{P}^\varepsilon_{n,r}$ by the projection $\pi : \mathcal{E}_r \to \widetilde{\mathcal{E}}_r$. With these preliminary definitions, we can now state the LDP result for $(\widetilde{\mathbb{P}}^\varepsilon_{n,r})_{n\in\mathbb{N}}$. The result is almost identical as Proposition 3.10.1, the only difference being the choice of truncation of the entries. Thus, the rate function is identical, and only the speed is different.

4.5.14 Proposition. *Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Then $(\widetilde{\mathbb{P}}^\varepsilon_{n,r})_{n\in\mathbb{N}}$ satisfies a LDP with speed $n^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$, and good rate function $I_{\varepsilon,r}$ defined for all $\tilde{A} \in \widetilde{\mathcal{E}}_r$ by*

$$I_{\varepsilon,r}\left(\tilde{A}\right) = \begin{cases} b\sum_{i\geq 1}|A_{i,i}|^\alpha + \frac{a}{2}\sum_{i\neq j}|A_{i,j}|^\alpha & \text{if } A \in \mathcal{D}_{\varepsilon,r}, \\ +\infty & \text{otherwise,} \end{cases}$$

where A is a representative of the equivalence class $\tilde{A}$ and

$$\mathcal{D}_{\varepsilon,r} = \left\{A \in \mathcal{E}_r : \forall i \leq j,\ A_{i,j} = 0 \text{ or } \varepsilon \leq |A_{i,j}| \leq \varepsilon^{-1}, \text{ and } A_{i,j}/|A_{i,j}| \in \operatorname{supp}(\nu_{i,j})\right\},$$

with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in definition 3.2.1.

We are now ready to use a contraction principle to prove that $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$ follows a LDP for any $\varepsilon > 0$. The use of the contraction principle is made possible by the fact that the push-forward of $\tilde{\mathbb{P}}^\varepsilon_{n,r}$ by the map $A \mapsto \mathrm{tr}A^p$ on $\cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)}$, are exponentially good approximations of $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$.

4.5.15 Proposition. *Let $\varepsilon > 0$. The sequence $(\mathrm{tr}_N\,(C^\varepsilon)^p)_{N\in\mathbb{N}}$ satisfies a large deviations principle of speed $N^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$, and good rate function J_ε defined for all $x \in \mathbb{R}$ by,*

$$K_\varepsilon(x) = \inf\left\{I_\varepsilon(A) : x = \mathrm{tr}A^p, A \in \cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)}\right\},$$

where

$$\forall A \in \cup_{n\in\mathbb{N}}\mathcal{H}_n^{(\beta)},\ I_\varepsilon(A) = \begin{cases} b\sum_{i\geq 1}|A_{i,i}|^\alpha + \frac{a}{2}\sum_{i\neq j}|A_{i,j}|^\alpha & \text{if } A \in \mathcal{D}_\varepsilon, \\ +\infty & \text{otherwise,}\end{cases} \tag{4.32}$$

where $\mathcal{D}_\varepsilon = \cup_{r\in\mathbb{N}}\mathcal{D}_{\varepsilon,r}$, with $\mathcal{D}_{\varepsilon,r}$ as in Proposition 3.10.1.

Proof. The proof will follow the lines of Proposition 3.10.3. Let $r \in \mathbb{N}$. We denote by f the function $\tilde{A} \in \tilde{\mathcal{E}}_r \mapsto \mathrm{tr}A^p$, with A a representative of $\tilde{A}$. As the trace is invariant by conjugation, f is well defined. We define the push-forward of $\tilde{P}^\varepsilon_{n,r}$ by the map f,

$$\nu_{n,r} = \tilde{P}^\varepsilon_{n,r} \circ f^{-1}.$$

Note that $\nu_{n,r}$ is the law of $\tau_n(C^\varepsilon)^p$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$. We will show that $(\nu_{n,r})_{n,r\in\mathbb{N}}$ are exponentially good approximations of $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$. Let $Y_{n,r}$ be random variable independent of C^ε, and distributed according to $\nu_{n,r}$. Let

$$Z_{n,r} = \tau_n\left(C^\varepsilon\right)^p \mathbb{1}_{C^\varepsilon\in\mathcal{E}_r} + Y_{n,r}\mathbb{1}_{C^\varepsilon\notin\mathcal{E}_r}.$$

Thus, $Z_{n,r}$ and $Y_{n,r}$ have the same law $\nu_{n,r}$. Furthermore, for any $t > 0$,

$$\mathbb{P}\left(|Z_{n,r} - \tau_n\left(C^\varepsilon\right)^p| > t\right) \leq \mathbb{P}\left(C^\varepsilon \notin \mathcal{E}_r\right).$$

By Proposition 4.5.7, we get

$$\lim_{n\to+\infty} n^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\left(|Z_{n,r} - \tau_n\left(C^\varepsilon\right)^p| > t\right) = -\infty,$$

which shows that $(\nu_{n,r})_{n,r\in\mathbb{N}}$ are exponentially good approximations of $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$.

For each $r \in \mathbb{N}$, the function f restricted to $\widetilde{\mathcal{E}}_r$ is continuous for the topology we equipped $\widetilde{\mathcal{E}}_r$. Note that as C^ε has entries bounded by $\varepsilon^{-1}n^{1/p}$, $\nu_{n,r}$ is compactly supported uniformly in n. Thus, $(\nu_{n,r})_{n\geq 1}$ is exponentially tight, the contraction principle (see [43, Theorem 4.2.1]) yields that $(\nu_{n,r})_{n\in\mathbb{N}}$ follows a LDP principle with speed $n^{\alpha(\frac{1}{2}+\frac{1}{p})}$ and good rate function $J_{\varepsilon,r}$ given by

$$K_{\varepsilon,r}(x) = \inf\left\{I_{\varepsilon,r}(\tilde{A}) : \tilde{A} \in \widetilde{\mathcal{E}}_r, x = f(\tilde{A})\right\},$$

where $I_{\varepsilon,r}$ is defined in Proposition 3.10.1. We can re-write this rate function as

$$K_{\varepsilon,r}(x) = \inf\left\{I_{\varepsilon}(A) : A \in \mathcal{E}_r,\ x = f(A)\right\},$$

where f denote as well the function $A \mapsto \operatorname{tr}(A)^p$ on $\cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)}$, and where I_ε is defined in (4.32). By [43, Theorem 4.2.16], we deduce that $(\tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$ satisfies a weak LDP with speed $n^{\alpha(\frac{1}{2}+\frac{1}{p})}$, and rate function K_ε defined by

$$\forall x \in \mathbb{R},\ K_\varepsilon(x) = \sup_{\delta>0} \liminf_{r\to+\infty} \inf_{|y-x|<\delta} K_{\varepsilon,r}(y).$$

As $K_{\varepsilon,r}$ is non-increasing in r, we have

$$K_\varepsilon(x) = \sup_{\delta>0} \inf_{r\in\mathbb{N}} \inf_{|y-x|<\delta} K_{\varepsilon,r}(y) = \sup_{\delta>0} \inf_{|y-x|<\delta} \inf_{r\in\mathbb{N}} K_{\varepsilon,r}(y).$$

Let Φ be the function defined by

$$\forall x \in \mathbb{R},\ \Phi(x) = \inf_{r\in\mathbb{N}} K_{\varepsilon,r}(x).$$

Thus,

$$K_\varepsilon(x) = \sup_{\delta>0} \inf_{|y-x|<\delta} \Phi(y).$$

We see that it suffices to show that Φ is lower semi-continuous to conclude that $K_\varepsilon = \Phi$. We will prove in fact that Φ has compact level sets.

Let $\tau > 0$. Let $x \in \mathbb{R}$, such that $\Phi(x) \le \tau$. Then

$$\Phi(x) = \left\{I_\varepsilon(A) : x = f(A), I_\varepsilon(A) \le 2\tau\right\}.$$

But for any $A \in \cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)}$ such that $I_\varepsilon(A) < +\infty$, we have

$$(b \wedge \frac{a}{2})\varepsilon^\alpha \operatorname{Card}\left\{(i,j) : A_{i,j} \neq 0\right\} \le I_\varepsilon(A).$$

Thus taking r such that $(b \wedge \frac{a}{2})\varepsilon^\alpha \le \tau$, we get

$$\begin{aligned}\Phi(x) &= \left\{I_\varepsilon(A) : x = f(A), I_\varepsilon(A) \le 2\tau, A \in \mathcal{E}_r\right\} \\ &= \left\{I_{\varepsilon,r}(\tilde{A}) : x = f(A), \tilde{A} \in \widetilde{\mathcal{E}}_r\right\}.\end{aligned}$$

Since f is continuous on $\widetilde{\mathcal{E}}_r$ and $I_{\varepsilon,r}$ is a good rate function, we have

$$\{x \in \mathbb{R} : \Phi(x) \le \tau\} = \left\{f(\tilde{A}) : I_{\varepsilon,r}(\tilde{A}) \le \tau, \tilde{A} \in \widetilde{\mathcal{E}}_r\right\}.$$

As f is continuous on $\mathcal{E}_r$, and $I_{\varepsilon,r}$ is a good rate function, we deduce that the τ-level sets of Φ are compact. Therefore $K_\varepsilon = \Phi$.

□

We are now ready to conclude the proof of Theorem 4.2.6

Proof of Theorem 4.2.6 . By Proposition 4.5.12, $(\langle \mu_{sc}, x^p \rangle + \tau_n(C^\varepsilon)^p)_{n\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations of $(\tau_n(X/\sqrt{n})^p)_{s\in\mathbb{N}}$. We deduce from Proposition 4.5.15 that for each $\varepsilon > 0$, the sequence $(\langle \mu_{sc}, x^p \rangle + \tau_n(C^\varepsilon)^p)_{n\in\mathbb{N}}$ satisfies a LDP with speed $n^{\alpha(\frac{1}{2}+\frac{1}{p})}$, and with good rate function ψ_ε defined by

$$\psi_\varepsilon(x) = \begin{cases} K_\varepsilon\left(x - C_{p/2}\right) & \text{if } p \text{ is even,} \\ K_\varepsilon(x) & \text{if } p \text{ is odd,} \end{cases}$$

where K_ε is as in Proposition 4.5.15. Since $(\tau_n(X/\sqrt{n})^p)_{n\geq 1}$ is exponentially tight by Proposition 4.5.3, we deduce from [43, Theorem 4.2.16] that $(\tau_n(X/\sqrt{n})^p)_{n\in\mathbb{N}}$ satisfies a LDP with speed $n^{\alpha(\frac{1}{2}+\frac{1}{p})}$ and rate function K_p defined by

$$\forall x \in \mathbb{R},\ K_p(x) = \sup_{\delta>0} \limsup_{\varepsilon\to 0} \inf_{|y-x|<\delta} \psi_\varepsilon(y).$$

Observe that for any $A \in \cup_{m\in\mathbb{N}} \mathcal{H}_m^{(\beta)}$, $I_\varepsilon(A)$ is non-decreasing in ε. Therefore, ψ_ε is non-decreasing in ε. Thus,

$$\forall x \in \mathbb{R},\ K_p(x) = \sup_{\delta>0} \inf_{\varepsilon>0} \inf_{|y-x|<\delta} \psi_\varepsilon(y). \tag{4.33}$$

Let

$$\forall x \in \mathbb{R},\ \Phi_p(x) = \begin{cases} \varphi_p\left(x - C_{p/2}\right) & \text{if } p \text{ is even,} \\ \varphi_p(x) & \text{if } p \text{ is odd,} \end{cases}$$

with

$$\varphi_p(x) = \inf\left\{W_\alpha(A) : x = \mathrm{tr} A^p, A \in \mathcal{D}\right\},$$

where W_α is defined, as in chapter 3, for any $A \in \cup_{m\geq 1} \mathcal{H}_m^{(\beta)}$, by

$$W_\alpha(A) = b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + a \sum_{i<j} |A_{i,j}|^\alpha,$$

and $\mathcal{D} = \{\cup_{m\in\mathbb{N}} \mathcal{H}_m^{(\beta)} : \forall i \leq j,\ A_{i,j} = 0 \text{ or and } A_{i,j}/|A_{i,j}| \in \mathrm{supp}(\nu_{i,j})\}$. With these notations we have,

$$K_p(x) = \sup_{\delta>0} \inf_{|x-y|<\delta} \Phi_p(y). \tag{4.34}$$

As for any $t > 0$, and $A \in \cup_{m\in\mathbb{N}} \mathcal{H}_m^{(\beta)}$, $I(tA) = t^\alpha I(A)$, and $\mathrm{tr}\,(tA)^p = t^p \mathrm{tr} A^p$, we have for p even,

$$\forall y \in \mathbb{R},\ \varphi_p(y) = \begin{cases} \varphi_p(1) y^{\alpha/p} & \text{if } y \geq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

and for p odd

$$\forall y \in \mathbb{R},\ \varphi_p(y) = \varphi_p(1)|y|^{\alpha/p}.$$

Therefore,

$$\forall x \in \mathbb{R},\ \Phi_p(x) = \begin{cases} \varphi_p(1)\left(x - C_{p/2}\right)^{\alpha/p} & \text{if } p \text{ is even,} \\ +\infty & \text{otherwise,} \end{cases}$$

and if p is odd

$$\forall x \in \mathbb{R},\ \Phi_p(x) = \varphi_p(1)|x|^{\alpha/p}.$$

This shows in particular that Φ_p is lower semi-continuous. From (4.34), we get finally $K_p = \Phi_p$.

□

4.5.10 Computation of $J_p(1)$

We show here that we can compute the constant c_p appearing in Theorem 4.2.6 when $\alpha \in (0,1]$ and p is even, and we give a lower bound and upper bound in the case where $\alpha \in (1,2)$ and p is even.

4.5.16 Theorem. *With the notations of Theorem 4.2.6, we have the following :*
(a). If p is even,

$$\min\left(b, \frac{a}{2}\right) \leq c_p \leq \min\left(b, 2^{-\alpha/p}a\right).$$

(b). If $\alpha \in (0,1]$ and p is even,

$$c_p = \min\left(b, 2^{-\alpha/p}a\right).$$

Proof. From the proof of Theorem 4.2.6, we know that

$$c_p = \inf\left\{W_\alpha(A) : 1 = \mathrm{tr}A^p, A \in \mathcal{D}\right\}, \tag{4.35}$$

where W_α is defined for any $A \in \cup_{m\geq 1}\mathcal{H}_m^{(\beta)}$, by

$$W_\alpha(A) = b\sum_{i=1}^{+\infty}|A_{i,i}|^\alpha + a\sum_{i<j}|A_{i,j}|^\alpha,$$

and $\mathcal{D} = \{\cup_{m\in\mathbb{N}}\mathcal{H}_m^{(\beta)} : \forall i \leq j,\ A_{i,j} = 0 \text{ or } A_{i,j}/|A_{i,j}| \in \mathrm{supp}(\nu_{i,j})\}$, with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in definition 3.2.1.

Note that

$$c_p \leq \min\left(W_\alpha(s), W_\alpha\begin{pmatrix} 0 & 2^{-1/p}e^{i\theta} \\ 2^{-1/p}e^{-i\theta} & 0 \end{pmatrix}\right),$$

where $s \in \mathrm{supp}(\nu_1)$, and $\theta \in \mathrm{supp}(\nu_2)$. Thus,

$$c_p \leq \min\left(b, 2^{-\alpha/p}a\right),$$

which proves the upper bound in cases (a) and (b).

On the other hand, we have

$$c_p \geq \inf \left\{ b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)}, 1 = \mathrm{tr} A^p \right\}$$
$$\geq \min\Big(b, \frac{a}{2}\Big) \inf \left\{ \sum_{i,j} |A_{i,j}|^\alpha : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)} : \mathrm{tr} A^p = 1 \right\}.$$

Since $\alpha \in (0,2)$, we know from [104, Theorem 3.32] that for any $A \in \mathcal{H}_m^{(\beta)}$,

$$\sum_{i,j} |A_{i,j}|^\alpha \geq \sum_{i=1}^{n} |\lambda_i|^\alpha, \tag{4.36}$$

where $\lambda_1, ..., \lambda_m$ are the eigenvalues of A. As $\alpha/p \leq 1$, we have

$$\sum_{i=1}^{m} |\lambda_i|^\alpha \geq \Big(\sum_{i=1}^{m} |\lambda_i|^p \Big)^{\alpha/p} = \Big(\mathrm{tr} |A|^p \Big)^{\alpha/p} \geq \Big| \mathrm{tr} A^p \Big|^{\alpha/p}.$$

Thus, if $\mathrm{tr} A^p = 1$, we have

$$\sum_{i,j} |A_{i,j}|^\alpha \geq 1.$$

We can deduce that

$$c_p \geq \min\Big(b, \frac{a}{2}\Big),$$

which proves the lower bound of case (b).

Assume now $\alpha \in (0,1)$ and p is even. If $A \in \mathcal{H}_m^{(\beta)}$ is such that $\mathrm{tr} A^p = 1$, then

$$\sup_{\mathrm{tr}|B|^q = 1} \mathrm{tr} AB = 1,$$

with $q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Thus, we can deduce that

$$\forall i \in \{1, ..., m\}, |A_{i,i}| \leq 1,\ \forall i,j \in \{1, ..., m\}, i \neq j, |A_{i,j}| \leq 2^{-1/p}.$$

Then,

$$c_p \geq \inf \Big\{ b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)}, 1 = \mathrm{tr} A^p \Big\} \tag{4.37}$$
$$\geq \min\Big(b, 2^{-\frac{\alpha}{p}} a\Big) \inf \Big\{ \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{1}{2} \sum_{i \neq j} |2^{\frac{1}{p}} A_{i,j}|^\alpha : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)}, 1 = \mathrm{tr} A^p \Big\}$$
$$\geq \min\Big(b, 2^{-\frac{\alpha}{p}} a\Big) \inf \Big\{ \sum_{i=1}^{+\infty} |A_{i,i}| + \frac{1}{2} \sum_{i \neq j} |2^{\frac{1}{p}} A_{i,j}| : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)}, 1 = \mathrm{tr} A^p \Big\},$$

where we used in the last inequality the fact the $|A_{i,i}| \leq 1$, and $|A_{i,j}| \leq 2^{-1/p}$ for any $i \neq j$. Thus,

$$c_p \geq \min\Big(b, 2^{-\frac{\alpha}{p}} a\Big) \inf \Big\{ \Big(1 - 2^{\frac{1}{p}-1}\Big) \sum_{i=1}^{+\infty} |A_{i,i}| + 2^{\frac{1}{p}-1} \sum_{i,j} |A_{i,j}| : A \in \cup_{m \in \mathbb{N}} \mathcal{H}_m^{(\beta)}, 1 = \mathrm{tr} A^p \Big\}.$$

Using again [104, Theorem 3.36], and the triangular inequality, we get

$$c_p \geq \min\left(b, 2^{-\frac{\alpha}{p}} a\right) \inf_{m\geq 1} \inf \left\{ \left(1 - 2^{\frac{1}{p}-1}\right) \left|\sum_{i=1}^{n} \lambda_i\right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i| : A \in \mathcal{H}_m^{(\beta)}, \sum_{i=1}^{m} \lambda_i^p = 1 \right\}.$$

Let $n \geq 1$. We consider the optimization problem

$$\inf \left\{ \left(1 - 2^{\frac{1}{p}-1}\right) \left|\sum_{i=1}^{n} \lambda_i\right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i| : A \in \mathcal{H}_n^{(\beta)}, \sum_{i=1}^{n} \lambda_i^p = 1 \right\}.$$

Denote for all $\lambda \in \mathbb{R}^n$,

$$\varphi(\lambda) = \left(1 - 2^{\frac{1}{p}-1}\right) \left|\sum_{i=1}^{n} \lambda_i\right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i|.$$

Compactness and continuity arguments show that the infimum is achieved at some $\lambda \in \mathbb{R}^n$. At the price of permuting the coordinates of λ, and taking the opposite of λ, which does not change the value of $\varphi(\lambda)$, we can assume that $\lambda = (\lambda_1, ..., \lambda_m, 0, ..., 0)$, with $\lambda_1 \neq 0, ..., \lambda_m \neq 0$ such that $\sum_{i=1}^m \lambda_i \geq 0$. Assume first that $\sum_{i=1}^m \lambda_i > 0$. The multipliers rule (see [40, Theorem 9.1]) yields that there is some $\gamma > 0$, such that for any $i \in \{1, ..., m\}$,

$$\left(1 - 2^{\frac{1}{p}-1}\right) + 2^{\frac{1}{p}-1} \mathrm{sg}(\lambda_i) = \gamma \lambda_i^{p-1}. \tag{4.38}$$

Multiplying the above inequality by λ_i, and summing over all $i \in \{1, ..., m\}$, we get

$$\gamma = \varphi(\lambda). \tag{4.39}$$

From (4.38), we have for all $\in \{1, ..., m\}$,

$$\lambda_i = \begin{cases} \gamma^{-\frac{1}{p-1}} & \text{if } \lambda_i > 0, \\ -\gamma^{-\frac{1}{p-1}} \left(2^{\frac{1}{p}} - 1\right)^{\frac{1}{p-1}} & \text{if } \lambda_i < 0. \end{cases}$$

Let k denote the number of positive λ_i's, and l the number of negative λ_i's. As $\sum_{i=1}^m \lambda_i > 0$, we have $k \geq 1$. Since $\sum_{i=1}^m \lambda_i^p = 1$, we have

$$\gamma^{\frac{p}{p-1}} = k + l\left(2^{\frac{1}{p}} - 1\right)^{\frac{p}{p-1}} \geq 1,$$

as $k \geq 1$. Thus, $\varphi(\lambda) \geq 1$.

Assume now that $\sum_{i=1}^m \lambda_i = 0$. Then the multipliers rule asserts that there are some $t \in [-1, 1]$ and γ, such that $(t, \gamma) \neq (0, 0)$, and for all $i \in \{1, ..., m\}$,

$$\left(1 - 2^{\frac{1}{p}-1}\right) t + 2^{\frac{1}{p}-1} \mathrm{sg}(\lambda_i) = \gamma \lambda_i^{p-1}.$$

At the price of changing λ to $-\lambda$, we can assume $t \geq 0$. As in the previous case, multiplying by λ_i in the above equation and summing over i, yields $\varphi(\lambda) = \gamma$. Note that since $\varphi(1, 0, ..., 0) = 1$, we can assume $\gamma \leq 1$. We can write for any $i \in \{1, ..., m\}$,

$$\lambda_i = \begin{cases} -\gamma^{-\frac{1}{p-1}} \left(2^{\frac{1}{p}-1} - \left(1 - 2^{\frac{1}{p}-1}\right)t\right)^{\frac{1}{p-1}} & \text{if } \lambda_i < 0, \\ \gamma^{-\frac{1}{p-1}} \left(2^{\frac{1}{p}-1} + \left(1 - 2^{\frac{1}{p}-1}\right)t\right)^{\frac{1}{p-1}} & \text{if } \lambda_i > 0. \end{cases}$$

Let k denotes the number of positive coordinates of λ, and by l the number of negative coordinates. As $\sum_{i=1}^m \lambda_i = 0$, we have $k, l \geq 1$, and

$$k\Big(2^{\frac{1}{p}-1} + \Big(1 - 2^{\frac{1}{p}-1}\Big)t\Big)^{\frac{1}{p-1}} = l\Big(2^{\frac{1}{p}-1} - \Big(1 - 2^{\frac{1}{p}-1}\Big)t\Big)^{\frac{1}{p-1}}.$$

But then,

$$\varphi(\lambda) = 2^{\frac{1}{p}} k\gamma^{-\frac{1}{p-1}} \Big(2^{\frac{1}{p}-1} + \Big(1 - 2^{\frac{1}{p}-1}\Big)t\Big)^{\frac{1}{p-1}} \geq 2^{\frac{1}{p}} 2^{-\frac{1}{p}} = 1,$$

as $\gamma \leq 1$. As $\varphi(1, 0, .., 0) = 1$, we can conclude

$$\inf \{\varphi(\lambda) : ||\lambda||_p = 1\} = 1.$$

This yields,

$$c_p \geq \min\Big(b, 2^{-\frac{\alpha}{p}} a\Big),$$

in the case where p is even.

$\square$

5. An isoperimetric approach to large deviations

5.1 Introduction

We investigated in the last Chapters 3 and 4 some instances of heavy-tail phenomena occurring in the large deviations of random matrices. Looking closer at the proofs we gave, in the setting of Wigner matrices without Gaussian tails for the largest eigenvalues and the traces of powers on one hand, and for traces of powers of Gaussian Wigner matrices on the other hand, we observe that the large deviations are created in a sense by additive perturbations (which are sparse matrices in the setting of Wigner matrices without Gaussian tails). We believe this mechanism of deviations can occur quite systematically in the large deviations governed by a heavy-tail phenomenon, and we would like to give in this chapter some elements in this direction.

The approach Borell and Ledoux (see [32], [33] and [44, Section 5], [69]) developped for the large deviations of Wiener chaoses will be the main framework in which we will work in this chapter. Indeed, the main feature which stands out of Borell and Ledoux's proof is that the large deviations of Wiener chaoses are due to translations by elements of the Cameron-Martin space. We believe this approach to be extremely fruitful and can shed a new light on heavy-tail phenomena appearing in the large deviations of certain models, where the large deviations are created also, in a sense, by translations. We already saw in Chapter 4 that the strategy put forward by Borell and Ledoux for the deviations of the Wiener chaos was very efficient to derive a LDP for the traces of Gaussian Wigner matrices. We will apply their strategy of proof beyond the Gaussian setting and we will propose a general large deviations result for a certain class of functionals $f_n : \mathbb{R}^n \to \mathcal{X}$, where $\mathcal{X}$ is some metric space, under the probability measure ν_α^n, where $\nu_\alpha = Y_\alpha^{-1} e^{-|x|^\alpha} dx$, for which the large deviations are governed by translations.

As an application of this result, we will retrieve the large deviations principles of different spectral functionals of the so-called Wigner matrices without Gaussian tail. In the more restricted setting where we assume that the entries have a density with respect to Lebesgue measure which is proportional to $e^{-c|x|^\alpha}$, with $c > 0$, and $\alpha \in (0, 2)$, the large deviations principles known for the spectral measure, the largest eigenvalue and traces of polynomials will fall in a unified way from our general large deviation result.

Another application of this result will consist in a large deviations principle for the last-passage time when the weights follow the law $\mu_\alpha = Z_\alpha^{-1} e^{-x^\alpha} dx$, for some $\alpha \in (0,1)$.

5.2 Main result

Let us present the main result of this chapter. For $\alpha > 0$, we denote by ν_α the probability measure on $\mathbb{R}$ with density $Y_\alpha^{-1} e^{-|x|^\alpha}$ with respect to Lebesgue measure, and ν_α^n its n-fold product measure on $\mathbb{R}^n$. Similarly, we define μ_α the probability measure on $\mathbb{R}^+$ with density $Z_\alpha^{-1} e^{-x^\alpha}$. We will denote for any $h \in \mathbb{R}^n$,

$$||h||_{\ell^\alpha} = \big(\sum_{i=1}^n |x_i|^\alpha \big)^{1/\alpha}.$$

The purpose of the general large deviations result we will present, is to identify a class of functionals $f_n : \mathbb{R}^n \to \mathcal{X}$, where $\mathcal{X}$ is some metric space, for which the large deviations are explained by translations. Let us describe first informally the assumptions we will make. Let X_n follow tha law ν_α^n. We will assume that $f_n(X_n)$ admits a kind of deterministic equivalent under additive deformations, given by a certain function F_n, that is,

$$f_n(X_n + v(n)^{1/\alpha} h_n) \simeq F_n(h_n), \tag{5.1}$$

in probability, for any sequence $h_n \in \mathbb{R}^n$, $||h_n||_{\ell^\alpha} < +\infty$, where $v(n)$ will eventually be the speed of deviations. It is convenient to think of $F_n(h_n)$ as a deterministic equivalent of $f_n(X_n + v(n)^{1/\alpha} h_n)$, where we took the large n limit on the variable X_n. Under this assumption, we will show that a large deviations lower bound for $f_n(X_n)$ at speed $v(n)$ holds with rate function,

$$J_\alpha = \sup_{\delta > 0} \limsup_{\substack{n \to +\infty \\ n \in N}} I_{n,\delta},$$

where

$$\forall x \in \mathcal{X},\ I_{n,\delta}(x) = \inf\{ ||h||_{\ell^\alpha}^\alpha : d(F_n(h), x) < \delta, h \in \mathbb{R}^n \}.$$

This rate function J_α can be interpreted by saying that to make a deviation around some $F_n(h_n)$, X_n needs to make a translation by $v(n)^{1/\alpha} h_n$, which one pays at the exponential scale $v(n)$ by $||h_n||_{\ell^\alpha}^\alpha$.

For the upper bound, we will further assume that for any $r > 0$, the deterministic equivalent (5.1) holds uniformly in $||h_n||_{\ell^\alpha} \leq r$. The upper bound will rely on some isoperimetric arguments, where we will need, excepted in the Gaussian case, to neglect the Euclidian enlargements appearing in the deviations inequality satisfied by ν_α^n. We thus make the assumption that f_n has a small, in expectation, local Lipschitz constant with respect to $|| \, ||_{\ell^2}$ when $\alpha < 2$. Finally, under some compactness property of F_n, we will prove that a large deviation upper bound holds for $f_n(X_n)$ with speed $v(n)$ and rate function,

$$I_\alpha = \sup_{\delta > 0} \inf_{n \in N} I_{n,\delta}.$$

Thus, if we moreover assume that the upper bound rate function I_α matches the lower rate function, we will get a full large deviations principle with speed $v(n)$. More precisely, we will prove the following result.

5.2.1 Theorem. *Let $(\mathcal{X}, d)$ be a metric space. Let $\alpha \in (0,2]$ and $N \subset \mathbb{N}$ an infinite subset. Let X_n be a random variable distributed according to ν_α^n. Let $f_n, F_n : \mathbb{R}^n \to \mathcal{X}$ be measurable functions. Let $(v(n))_{n\in N}$ be a sequence going to $+\infty$. Define for $\delta > 0$ and $n \in N$, the function*

$$\forall x \in \mathcal{X},\ I_{n,\delta}(x) = \inf\{||h||_{\ell^\alpha}^\alpha : d(F_n(h), x) < \delta, h \in \mathbb{R}^n\}.$$

We set

$$\forall x \in \mathcal{X},\ I_\alpha(x) = \sup_{\delta>0} \inf_{n\in N} I_{n,\delta}(x). \tag{5.2}$$

We assume:
(i).(Uniform deterministic equivalent). For any $r > 0$,

$$\sup_{h_n \in rB_{\ell^\alpha}} d\big(f_n(X_n + v(n)^{1/\alpha} h_n), F_n(h_n)\big) \underset{\substack{n\to+\infty \\ n\in N}}{\longrightarrow} 0,$$

in probability.
(ii).(Control of the Lipschitz constant). If $\alpha < 2$, for any $\delta > 0$ and $r > 0$, there is a sequence $t_\delta(n)$ such that,

$$\mathbb{E} \sup_{||h||_{\ell^2} \leq t_\delta(n)} \mathcal{L}_n(h) \leq \delta,$$

with

$$\mathcal{L}_n(h) = \sup_{X_n + rv(n)^{1/\alpha} B_{\ell^\alpha}} d\big(f_n(x+h), f_n(x)\big),$$

satisfying,

$$(\log n)^{\alpha/2} = o(\log \frac{t_\delta(n)^2}{v(n)}) \text{ if } \alpha \neq 1, \text{ or } v(n) = o(t_\delta(n)^2) \text{ if } \alpha = 1.$$

(iii).(Compactness). For any $r > 0$, $\cup_{n\in N} F_n(rB_{\ell^\alpha})$ is relatively compact.
(iv).(Upper bound = lower bound). For any $x \in \mathcal{X}$,

$$I_\alpha(x) = \sup_{\delta>0} \limsup_{\substack{n\to+\infty \\ n\in N}} I_{n,\delta}(x). \tag{5.3}$$

Then $(f_n(X_n))_{n\in N}$ satisfies a LDP with speed $v(n)$ and good rate function I_α.

Let us make some remarks on the assumptions of this theorem.

5.2.2 Remarks. (a). We will prove that under the assumption that for any sequence $h_n \in \mathbb{R}^n$, $n \in N$, such that $\sup_n ||h_n||_{\ell^\alpha} < +\infty$,

$$d(f_n(X_n + v(n)^{1/\alpha} h_n), F_n(h_n)) \underset{\substack{n\to+\infty \\ n\in N}}{\longrightarrow} 0, \tag{5.4}$$

in probability, the lower bound of the LDP holds with the rate function (5.3).

(b). The assumption (i) that the approximation (5.4) holds uniformly in $h_n \in rB_{\ell^\alpha}$ is crucial for deriving the upper bound of the LDP with rate function (5.2), and is one of the most constraining assumptions of Theorem 5.2.1. In the applications we develop for $\alpha < 2$, this is proven by some concentration inequality and chaining arguments, which can be carried out successfully due to the "sparsity" of the ball B_{ℓ^α}.

(c). The formulation of assumption (ii) on the Lipschitz constant of f_n is specially designed to include polynomial functionals f_n, as the trace of a polynomial of random matrices. In other words, it says that the "local" Lipschitz constant of f_n, is small enough uniformly on the set $X_n + rv(n)^{1/\alpha}B_{\ell^\alpha}$. Note that when f_n is $L_2(n)$-Lipschitz with respect to $|| \, ||_{\ell^2}$, a sufficient condition for assumption (ii) to be fulfilled is

$$(\log n)^{\alpha/2} = o\big(\log \frac{1}{L_2(n)^2 v(n)}\big). \tag{5.5}$$

For $\alpha < 2$, this assumption enables us to neglect the Euclidean ball in the deviations inequality for ν_α^n. As we will see in the proof, it ensures that the deviations of $f_n(X_n)$ are explained by a heavy-tail phenomenon. For example, it fails to hold for empirical means under ν_α^n when $\alpha \in [1, 2)$.

(d). The compactness assumption of (iii) is made to ensure that I_α is a good rate function. As one can observe in the proof, without it, the upper bound of the LDP holds only for compact sets.

(e). The rate function I_α can be simplified in certain cases. Define the function $\tilde{I}_\alpha$ by,

$$\forall x \in \mathcal{X}, \;\; \tilde{I}_\alpha(x) = \inf_n \{ ||h||_{\ell^\alpha}^\alpha : x = F_n(h), \;\; h \in \mathbb{R}^n \}.$$

One can see that,

$$I_\alpha = \sup_{\delta > 0} \inf_{B(x,\delta)} \tilde{I}_\alpha.$$

Thus if $\tilde{I}_\alpha$ is lower semi-continuous, then $I_\alpha = \tilde{I}_\alpha$.

The proof is greatly inspired from the ideas and the framework developed by Borell and Ledoux in [32], [33] and [44, Section 5, Theorem 5.1] for the large deviations for Wiener chaoses. To make a parallel their approach, for the lower bound, we replace the use of the Cameron-Martin formula, used in the context of abstract Wiener space, with a lower bound estimate of the probability of translated events, that is,

$$\liminf_{\substack{n \to +\infty \\ n \in N}} \frac{1}{v(n)} \log \nu_\alpha^n (E + v(n)^{1/\alpha} h_n) \geq - \limsup_{\substack{n \to +\infty \\ n \in N}} c_\alpha(h_n), \tag{5.6}$$

for a given sequence $h_n \in \mathbb{R}^n$, and subsets E such that $\liminf_n \nu_\alpha^n(E) > 0$. In the Gaussian case $\alpha = 2$, this estimates is given by the translation formula of the Gaussian measure, where $c_\alpha(h) = ||h||_{\ell^2}^2$. When $\alpha < 2$, one can mimic the Gaussian case to get such an estimate (5.6) with $c_\alpha(h) = ||h||_{\ell^\alpha}^\alpha$, whereas when $\alpha > 2$, we believe that there is a competition between the speed and the dimension which is not workable in the applications.

Whereas the Gaussian isoperimetric inequality is used in the proof of the upper bound deviations of Wiener chaoses, ours will rely on sharp large deviations inequalities for ν_α^n with respect to the weight function c_α, that is

$$\limsup_{\substack{n\to+\infty \\ n\in N}} \frac{1}{v(n)} \log \nu_\alpha^n(x \notin E + \{c_\alpha \leq rv(n)\}) \leq -r, \tag{5.7}$$

for some "large enough" subsets E. We will show that we can take $c_\alpha = ||h||_{\ell^\alpha}^\alpha$, which together with (5.6) will allow us to make the upper and lower bound match. In the Gaussian case, this is due to the Gaussian isoperimetric inequality, whereas when $\alpha < 2$, we will have to call for sharp inf-convolution inequalities for ν_α^n. This is in particular where assumption (ii) plays its role since it enables us, when $\alpha < 2$, to neglect the Euclidean balls which come naturally in the deviation inequality of ν_α^n, and consider subsets E which are indeed large enough.

These two estimates (5.6) and (5.7) are behind the limitation in Theorem 5.2.1 to the probability measures ν_α^n for $\alpha \in (0,2]$. For example, if one replace the measure ν_α by the probability measure on $\mathbb{R}_+$ with density $Z_\alpha^{-1}e^{-x^\alpha}$, one can show that (5.6) holds provided h_n has all its coordinates non-negative (and $n = o(v(n))$ if $\alpha > 1$). But then, we will have to prove (5.7) with $c_\alpha(h) = ||h||_{\ell^\alpha}^\alpha$ if the coordinates of h are non-negative, and $+\infty$ otherwise, which we do not know how to obtain.

This said, we can give a version of Theorem 5.2.1 for the probability measure μ_α, with density $Z_\alpha^{-1}e^{-x^\alpha}\mathbb{1}_{x\geq 0}$, which will be sufficient to prove a LDP result for the last-passage time.

5.2.3 Theorem. *Let $\alpha \in (0,1]$ and $N \subset \mathbb{N}$ an infinite subset. Let X_n be a random variable distributed according to μ_α^n. Let $f_n, F_n : \mathbb{R}^n \to \mathbb{R}$ be measurable functions. Let $(v(n))_{n\in N}$ be a sequence going to $+\infty$. Define I_α as in (5.2), and for $\delta > 0$ and $n \in N$,*

$$I_{n,\delta}^+(x) = \inf\left\{||h||_{\ell^\alpha}^\alpha : d(F_n(h), x) < \delta, h \in \mathbb{R}_+^n\right\}.$$

Assume $(i)-(ii)-(iii)$ from Theorem 5.2.1, and,
(iv)': For any $x \in \mathcal{X}$,

$$I_\alpha(x) = \sup_{\delta>0} \limsup_{\substack{n\to+\infty \\ n\in N}} I_{n,\delta}^+(x).$$

Then $(f_n(X_n))_{n\in N}$ satisfies a LDP with speed $v(n)$ and good rate function I_α.

5.2.4 Remark. We only state this result for $\alpha \in (0,1]$ because for $\alpha > 1$, we know how to get the lower bound (5.6) for a sequence $h_n \in \mathbb{R}_+^n$ only under the additional assumption on the speed that $n = o(v(n))$. But this condition and the requirement (ii) cannot be met simultaneously the application we will present.

5.2.1 Applications to Wigner matrices

We present now the applications of Theorem 5.2.1 to Wigner matrices. We denote by $\mathcal{H}_n^{(\beta)}$ the set of Hermitian matrices when $\beta = 2$ and symmetric matrices when $\beta = 1$ of size n. We define $\mathcal{S}_\alpha$ the class of Wigner matrices whose law is of density

$Z_{W_\alpha}^{-1}e^{-W_\alpha}$ with respect to the Lebesgue measure $\ell_n^{(\beta)}$ on $\mathcal{H}_n^{(\beta)}$, where

$$\forall A \in \mathcal{H}_n^{(\beta)},\ W_\alpha(A) = b\sum_i |A_{i,i}|^\alpha + \sum_{i<j}\Big(a_1|\Re A_{i,j}|^\alpha + a_2|\Im A_{i,j}|^\alpha\Big), \tag{5.8}$$

for some $b, a_1, a_2 \in (0, +\infty)$, and where Z_{W_α} is the normalizing constant.

We recall that we denote by μ_A and λ_A, respectively the spectral measure and the largest eigenvalue of a matrix $A \in \mathcal{H}_n^{(\beta)}$.

As a consequence of Theorem 5.2.1, we have the following large deviations principles, originally proven in [29] in the case of the spectral measure and in [6] for the largest eigenvalue (see Chapter 3).

5.2.5 Theorem. *Let $\alpha \in (0,2)$. Assume that $X \in \mathcal{S}_\alpha$, and $\mathbb{E}|X_{1,2}|^2 = 1$. $(\mu_{X/\sqrt{n}})_{n\in\mathbb{N}}$ follows a LDP with respect to the weak topology with speed $n^{1+\alpha/2}$, and good rate function I_α, defined for any probability measure μ on $\mathbb{R}$ by,*

$$I_\alpha(\mu) = \sup_{\delta>0}\inf_{n\in\mathbb{N}}\{W_\alpha(A) : A \in \mathcal{H}_n^{(\beta)},\ d(\mu, \mu_{sc}\boxplus\mu_{n^{1/\alpha}A}) < \delta\},$$

where d is a distance compatible with the weak topology.

5.2.6 Remark. In [29], the rate function I_α is computed explicitly for symmetric probability measures ν, for which we have

$$I_\alpha(\nu) = \min\big(b, \frac{a}{2}\big)\int |x|^\alpha d\nu(x).$$

5.2.7 Theorem. *Let $\alpha \in (0,2)$. Assume that $X \in \mathcal{S}_\alpha$, and $\mathbb{E}|X_{1,2}|^2 = 1$. $(\lambda_{X/\sqrt{n}})_{n\in\mathbb{N}}$ follows a LDP with speed $n^{\alpha/2}$ and good rate function J_α, defined for any $x \in \mathbb{R}$ by,*

$$J_\alpha(x) = \begin{cases} c g_{\mu_{sc}}(x)^{-\alpha} & \text{if } x \geq 2,\\ 0 & \text{if } x = 2,\\ +\infty & \text{if } x < 2,\end{cases}$$

with

$$c = \inf\big\{W_\alpha(A) : A \in \cup_{n\in\mathbb{N}}\mathcal{H}_n^{(\beta)},\ \lambda_A = 1\big\}.$$

5.2.8 Remark. The constant c can be computed explicitly in the case where the entries are real, as we saw in Chapter 3 at section 3.11.

Concerning the deviations of normalized traces of polynomials in independent matrices in the class $\mathcal{S}_\alpha$, with $\alpha \in (0,2]$ we have the following result.

5.2.9 Theorem. *Let $\alpha \in (0,2]$ and $p \in \mathbb{N}$, $p > \alpha$. Assume $\boldsymbol{X} = (X_1, ..., X_p)$ is a collection of independent Wigner matrices in the class $\mathcal{S}_\alpha$, such that for $M \in \{X_1, ..., X_p\}$, $\mathbb{E}|M_{1,2}|^2 = 1$. We assume that X_i is distributed according to $Z_{W_\alpha}^{-1}e^{-W_{\alpha,i}}d\ell_n^{(\beta)}$, where $W_{\alpha,i}$ is of the form (5.8). Let $P \in \mathbb{C}\langle\boldsymbol{X}\rangle$ be a non-commutative polynomial of total degree d. We denote by τ_n the state $\frac{1}{n}\mathrm{tr}$ on $\mathcal{H}_n^{(\beta)}$. The sequence*

$$\tau_n[P(\boldsymbol{X}/\sqrt{n})]$$

satisfies a LDP with speed $n^{\alpha\left(\frac{1}{2}+\frac{1}{d}\right)}$ *and good rate function* K_α*, defined for all* $x \in \mathbb{R}$ *by*

$$K_\alpha(x) = \begin{cases} c_1\big(x - \tau(P(\boldsymbol{s}))\big)^{\frac{\alpha}{d}} & \text{if } x > \tau(P(\boldsymbol{s})), \\ 0 & \text{if } x = \tau(P(\boldsymbol{s})), \\ c_{-1}\big|x - \tau(P(\boldsymbol{s}))\big|^{\frac{\alpha}{d}} & \text{if } x < \tau(P(\boldsymbol{s})), \end{cases}$$

where for any $\sigma \in \{-1, 1\}$,

$$c_\sigma = \inf\left\{W_\alpha(\boldsymbol{H}) : \boldsymbol{H} \in \cup_{n\in\mathbb{N}}(\mathcal{H}_n^{(\beta)})^p, \sigma = \mathrm{tr}P_d(\boldsymbol{H})\right\} \in [0, +\infty],$$

where $W_\alpha(\boldsymbol{H}) = \sum_{i=1}^p W_{\alpha,i}(H_i)$*,* P_d *is the homogeneous part of degree* d *of* P*, and* $\boldsymbol{s} = (s_1, ..., s_p)$ *is a free family of* p *semi-circular variables in a non-commutative probability space* $(\mathcal{A}, \tau)$*.*

5.2.10 Remark. Unlike the previous results on deviations of the spectral measure and the largest eigenvalue, this one allows us to consider Gaussian matrices. As we will see in the proof, the mechanism of deviations of traces of polynomials is the same in both cases $\alpha \in (0, 2)$, and $\alpha = 2$. This is essentially due to the fact that still in the Gaussian case there is a heavy-tail phenomena which appears when the degree of the polynomial is strictly greater than 2 since there is no exponential moments.

This large deviations principle is an extension, although in a more restricted setting, of the large deviations principle proven in Chapter 4, in the case where $p = 1$ and $P = X^d$ for some $d \geq 3$, for Gaussian matrices and Wigner matrices without Gaussian tails.

5.2.2 Application to last-passage percolation

Let $d \in \mathbb{N}$, $d \geq 2$. We denote by $\mathbb{Z}_+^d$ the subset of vectors of $\mathbb{Z}^d$ with non-negative coordinates. Let $(X_v)_{v\in\mathbb{Z}_+^d}$ be a collection of weights. We will call a *directed path* a path in which at each step, one coordinate is increased by 1. For $v_1, v_2 \in \mathbb{Z}_+^d$, we denote by $\Pi(v_1, v_2)$ the set of directed paths from v_1 to v_2. We will identify a path with the set of its vertices. We define the *last-passage time* $T_{v_1,v_2}(X)$, by

$$T_{v_1,v_2}(X) = \sup_{\pi\in\Pi(v_1,v_2)} \sum_{v\in\pi} X_v,$$

We know by a work of Martin [75], that if the weights X_v are i.i.d random variables with common distribution function F satisfying,

$$\int_0^{+\infty} (1 - F(t))^{1/d} dt < +\infty, \tag{5.9}$$

then for any $v \in \mathbb{R}_+^d$,

$$\frac{1}{n}\mathbb{E}T_{0,\lfloor nv\rfloor}(X) \underset{n\to+\infty}{\longrightarrow} g(v), \tag{5.10}$$

where g is a continuous function on $\mathbb{R}_+^d$.

As an application of Theorem 5.2.3, we will get the following LDP for the last-passage time.

5.2.11 Theorem. *Let $\alpha \in (0,1)$. For any $n \in \mathbb{N}$, we set $T(X) = T_{0,(n,\dots,n)}(X)$. Let $(X_v)_{v\in\mathbb{Z}_+^d}$ be a family of i.i.d random variables distributed according to μ_α. The sequence $T(X)/n$ satisfies a LDP with speed n^α and good rate function L_α, defined by*

$$L_\alpha(x) = \begin{cases} (x - g(1,\dots,1))^\alpha & \text{if } x \geq g(1,\dots,1), \\ +\infty & \text{otherwise.} \end{cases}$$

5.2.3 Organization of the chapter

The chapter is organized as followed. In the section 5.3, we prove some inf-convolution inequalities for ν_α^n. As the large deviations of our functional f_n are governed by translates, we will need some sharp deviations inequality with respect to the metric $|| \,||_{\ell^\alpha}$ (or $|| \,||_{\ell^\alpha}^\alpha$ when $\alpha < 1$). We will provide a family of weights $W_{\alpha,\varepsilon}$ which captures the asymptotics of the tail distribution of ν_α^n, that is, behaving like $||x||_{\ell^\alpha}^\alpha$ when $||x||_\infty \gg 1$. This will be done by transporting and tensoring the family of optimal weights known for the exponential law due to Talagrand [95].

In the section 5.4, we give a proof of Theorem 5.2.1. The upper bound relies on Proposition 5.4.1 which gives a large deviations sharp upper bound for ν_α^n with respect to the metric $|| \,||_{\ell^\alpha}$ using the inf-convolution inequalities proven in section 5.3. The lower bound is given by Proposition 5.4.4 which estimates at the exponential scale $v(n)$ the probability, under ν_α^n, of an event translated by some element $v(n)^{1/\alpha}h_n$.

The rest of the chapter is devoted to applications to Wigner matrices and the last-passage time.

We begin by proving in section 5.5 uniform deterministic equivalents for the spectral measure, largest eigenvalue and traces of non-commutative polynomials of deformed Wigner matrices. To make the equivalents for the spectral measure and largest eigenvalue of hold uniformly in the case of Wigner matrices in the class $\mathcal{S}_\alpha$ for $\alpha < 2$, we make use of the concentration inequalities we proved in Chapter 2, and perform a classical chaining argument.

In section 5.6, we provide a deterministic equivalent for the last-passage time under deformations of the weights. We use the same arguments as for the case of Wigner matrices to make this equivalent to hold uniformly, making use in particular of the deviation inequality we proved in Chapter 2 at section 2.2.4 for ν_α^n, and using again a chaining argument.

In section 5.7, we apply Theorem 5.2.1 in the setting of Wigner matrices in the class $\mathcal{S}_\alpha$, to the spectral measure, the largest eigenvalue (for $\alpha \in (0,2)$) and to traces of non-commutative polynomials (for $\alpha \in (0,2]$). Making use of the uniform deterministic equivalents we proved in section 5.5, we give a proof of Theorems 5.2.5, 5.2.7, and 5.2.9. Finally we prove in section 5.8, Theorem 5.2.11 by applying Theorem 5.2.3 and using again the uniform deterministic equivalent we proved in section 5.6.

5.3 Inf-convolution inequalities for ν_α^n

Let ν be a probability measure on $\mathbb{R}^n$, and let w be a measurable function on $\mathbb{R}^n$ taking non-negative values. Following Maurey (see [77]), we will say that (ν, w) satisfies the τ-property if for any non-negative measurable function f on $\mathbb{R}^n$,

$$\Big(\int e^{f\square w} d\nu\Big)\Big(\int e^{-f} d\nu\Big) \leq 1, \tag{5.11}$$

where $\square$ denotes the inf-convolution, that is,

$$\forall x \in \mathbb{R}^n,\ f\square w(x) = \inf_{y\in\mathbb{R}^n} \{f(y) + w(y-x)\}.$$

The τ-property is closely linked to transportation-cost inequalities. By the Kantorovitch duality (see [99, Theorem 5.10]), and the duality of the entropy (see [43, Lemma 6.2.13]), it is known that under mild assumptions of w that the following general inf-convolution inequality,

$$\int e^{f\square w} d\nu \leq e^{\int f d\nu}, \tag{5.12}$$

satisfied for any non-negative measurable functions f is equivalent to the following transportation-cost inequality: for any μ probability measure on $\mathbb{R}^n$,

$$\mathcal{W}_w(\mu,\nu) \leq D(\mu||\nu), \tag{5.13}$$

where $D(\mu||\nu)$ is the Kullback-Leibler divergence, and

$$\mathcal{W}_w(\mu,\nu) = \inf\Big\{\int w(x-y)d\pi(x,y) : \pi \text{ has marginals } \mu \text{ and } \nu\Big\}. \tag{5.14}$$

In particular, under the assumption that w is upper semi-continuous, Kantorovitch duality is valid by [99, Theorem 5.10], so that the equivalence above between (5.12) and (5.14) holds.

One can observe that if (ν, w) satisfies the τ-property, then by Jensen's inequality, it satisfies also the general inf-convolution inequality (5.12), and therefore ν satisfies the transportation-cost inequality (5.13) with cost function w.

Conversely, according to [54, Proposition 4.13], if ν satisfies the transportation-cost inequality (5.13) with cost function w, then $(\nu, w\square w)$ satisfies the τ-property. If moreover w is sub-additive, then (ν, w) satisfies the τ-property, whereas if w is convex, then $(\nu, 2w(./2))$ satisfies the τ-property.

More importantly for us, the τ-property yields deviations bounds with respect to enlargements by the weight w. We know from [77, Lemma 4], that if (ν, w) satisfies the τ-property, then for any Borel subset A of $\mathbb{R}^n$, and any $t > 0$,

$$\nu\big(x \notin A + \{w \leq r\}\big) \leq \frac{e^{-r}}{\nu(A)}. \tag{5.15}$$

We define another form of inf-convolution inequality, designed to enable us to get the best constants in our weight functions, (and also to deal with the measure

ν_α^n when $\alpha \in (0,1)$), which we will call the *truncated τ-property*. More precisely, we will say that a measure ν on $\mathbb{R}^n$ with the weight function w, satisfies the A_0-*truncated τ-property*, where A_0 is a measurable subset of $\mathbb{R}^n$, if (5.11) is true for any non-negative measurable function f such that $f = +\infty$ on A_0^c.

This A_0-truncated τ-property yields a deviation inequality with respect to enlargement by the weight w of the following form. For any Borel subset A of $\mathbb{R}^n$ such that $\nu(A) > 0$, and any $r > 0$,

$$\nu\big(x \notin A + \{w \le r\}\big) \le \frac{e^{-r}}{\nu(A \cap A_0)}. \tag{5.16}$$

The goal of this section is to find, for the measure ν_α^n, for $\alpha \in (0,2)$, a family of weights $W_{\alpha,\varepsilon}$ for which a truncated τ-property is satisfied, and which captures the asymptotics of the tail distribution of ν_α^n. More precisely, we will prove the following proposition.

5.3.1 Proposition. *Let $\alpha > 0$. If $\alpha = 1$, then for any $\varepsilon < 1/2$, $(\nu_1^n, W_{1,\varepsilon})$ satisfies the τ-property with*

$$W_{1,\varepsilon}(x) = \sum_{i=1}^n w_\varepsilon(x_i),$$

where

$$w_\varepsilon(t) = \begin{cases} \frac{\varepsilon e^{-1/\varepsilon} t^2}{8} & \text{if } |t| \le 2/\varepsilon^2, \\ (1-2\varepsilon)|t| & \text{if } |t| > 2/\varepsilon^2. \end{cases}$$

If $\alpha \neq 1$, there are some constants $\kappa > 0$ and $\varepsilon_0 \in (0,1)$ such that for $\varepsilon \in (0,\varepsilon_0)$ and $m \ge 1$, $(\nu_\alpha^n, W_{\alpha,\varepsilon}^{(m)})$ satisfies the mB_{ℓ^∞}-truncated τ-property, where

$$\forall x \in \mathbb{R}^n,\ W_{\alpha,\varepsilon}^{(m)}(x) = \sum_{i=1}^n w_{\alpha,\varepsilon}^{(m)}(x_i), \tag{5.17}$$

with

$$w_{\alpha,\varepsilon}^{(m)}(t) = \begin{cases} \kappa^{-1} e^{-(\frac{m}{\varepsilon})^{\alpha/2}} t^2 & \text{if } |t| \le m\varepsilon^{-1}, \\ (1-\kappa\varepsilon^{(\alpha/2)\wedge 1})|t|^\alpha & \text{if } |t| > m\varepsilon^{-1}. \end{cases}$$

The rest of this section will be devoted to proving the above proposition. We will reduce the problem in a first phase to the one-dimensional case, and to an estimation of the monotone rearrangement of ν_1 onto ν_α.

As the usual τ-property (see [77, Lemma 1]), the truncated version of the τ-property tensorizes in the following way.

5.3.2 Lemma. *Let ν_i be a probability measure defined on some measurable space $\mathcal{X}_i$, A_i be some measurable subset of $\mathcal{X}_i$ and $w_i : \mathcal{X}_i \to \mathbb{R}_+$ be a measurable function, for $i = 1, 2$.*

If (ν_i, w_i) satisfies the A_i-truncated τ-property for $i = 1, 2$, then $(\nu_1 \otimes \nu_2, w)$ satisfies the $A_1 \times A_2$-truncated τ-property with

$$\forall (x,y) \in \mathcal{X}_1 \times \mathcal{X}_2,\ w(x,y) = w_1(x) + w_2(y).$$

Since we are dealing with the product measure ν_α^n, we can focus on studying the τ-property for the one-dimensional marginal ν_α.

For the exponential measure, we have the following result due to Talagrand, which gives a family of optimal weights c_λ.

5.3.3 Proposition ([94, Theorem 1.2])**.** *Let $\lambda \in (0,1)$. Define the weight function c_λ for any $x \in \mathbb{R}$ by,*

$$c_\lambda(x) = \big(\frac{1}{\lambda} - 1\big)\big(e^{-\lambda|x|} - 1 + \lambda|x|\big).$$

For any $\lambda \in (0,1)$, ν_1 satisfies a transportation-cost inequality (5.13) with cost function c_λ.

Note that, $c_\lambda(x) \sim_{\pm\infty} (1-\lambda)|x|$. Thus, when $\lambda \ll 1$, c_λ captures the exact asymptotics of the tail distribution of the exponential law.

For technical reasons, we prefer to work with a different family of weights than the one defined in Proposition 5.3.3. In the following corollary, we reformulate Talagrand's result for the symmetric exponential measure ν_1.

5.3.4 Corollary. *Let $\varepsilon > 0$. We define the weight function w_ε, for any $t \in \mathbb{R}$, by*

$$w_\varepsilon(t) = \begin{cases} \frac{\varepsilon e^{-1/\varepsilon} t^2}{8} & \text{if } |t| \leq 2/\varepsilon^2, \\ (1-2\varepsilon)|t| & \text{if } |t| > 2/\varepsilon^2. \end{cases}$$

For any $\varepsilon \in (0,1/2)$, (ν_1, w_ε) satisfies the τ-property. As a consequence, $(\nu_1^n, W_{1,\varepsilon})$ satisfies the τ-property, with $W_{1,\varepsilon}$ defined in Proposition 5.3.1.

Proof. As c_λ is a convex function, we know by [54, Proposition 4.13] that $(\nu_1, 2c_\lambda(./2))$ satisfies the τ-property. To prove Corollary 5.3.4, it suffices to prove that $w_\varepsilon \leq 2c_\varepsilon(./2)$ for any $\varepsilon \in (0,1/2)$. Since both functions are symmetric, it is sufficient to prove the inequality on $\mathbb{R}_+$. For any $0 \leq t \leq 1/\varepsilon^2$, we have by Taylor's formula

$$e^{-\varepsilon t} - 1 + \varepsilon t = \varepsilon^2 e^{-\varepsilon y}\frac{t^2}{2},$$

for some $y \in [0,t]$. If $t \leq 2/\varepsilon^2$ and $\varepsilon \leq 1/2$, we get

$$2c_\varepsilon(t/2) \geq \varepsilon\big(1-\varepsilon\big)e^{-1/\varepsilon}\frac{t^2}{4} \geq w_\varepsilon(t).$$

If $t \geq 1/\varepsilon^2$, we have

$$c_\varepsilon(t) \geq \big(\frac{1}{\varepsilon} - 1\big)\big(-1+\varepsilon t\big) \geq (1-\varepsilon)t - \frac{1}{\varepsilon} \geq (1-2\varepsilon)t.$$

Thus, $2c_\varepsilon(t/2) \geq (1-2\varepsilon)t$ for $t \geq 2/\varepsilon^2$.

After tensorization (see [77, Lemma 1]), we obtain that $(\nu_1^n, W_{1,\varepsilon})$ satisfies the τ-property with $W_{1,\varepsilon}$ defined in Proposition 5.3.1.

□

For $\alpha \neq 1$, the general strategy is to transport this τ-property of the symmetric exponential law to obtain a τ-property for ν_α. In the next lemma, we describe how to get a τ-property using transport. It extends in our setting of truncated τ-property, a result of Maurey [77, Lemma 2].

5.3.5 Lemma. *Let A be a Borel subset of $\mathbb{R}^n$. Let μ be a probability measure on $\mathbb{R}^n$ and let $\psi : \mathbb{R}^n \to \mathbb{R}^n$ be a bijective measurable map. Assume (μ, w) satisfies the τ-property. Let A be a Borel subset of $\mathbb{R}^n$ and let $\tilde{w}$ be a weight function such that,*

$$\forall x \in \mathbb{R}^n, y \in A,\ \tilde{w}(x-y) \leq w\big(\psi^{-1}(x) - \psi^{-1}(y)\big).$$

Then, $(\mu \circ \psi^{-1}, \tilde{w})$ satisfies the A-truncated τ-property.

Proof. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable non-negative function being $+\infty$ on A^c. Applying the τ-property of (μ, w) to $f \circ \psi$, we get

$$\Big(\int e^{f\circ\psi \square w} d\mu\Big)\Big(\int e^{-f\circ\psi} d\mu\Big) \leq 1.$$

But, as ψ is a bijection and $f = +\infty$ on A^c,

$$\begin{aligned} f \circ \psi \square w(\psi^{-1}(x)) &= \inf_{y\in\mathbb{R}^n} \{f(y) + w(\psi^{-1}(x) - \psi^{-1}(y))\} \\ &= \inf_{y\in A} \{f(y) + w(\psi^{-1}(x) - \psi^{-1}(y))\}. \end{aligned}$$

From the assumption on $\tilde{w}$, we deduce

$$f \circ \psi \square w(\psi^{-1}(x)) \geq \inf_{y\in A} \{f(y) + w(x-y)\} \geq f \square w(x).$$

Therefore,

$$\Big(\int e^{f\square w} d\mu \circ \psi^{-1}\Big)\Big(\int e^{-f} d\mu \circ \psi^{-1}\Big) \leq 1.$$

□

In particular, in the one-dimensional case, if (μ, w) satisfies the τ-property and if w is even and non-decreasing on $\mathbb{R}_+$, then $\mu \circ \psi^{-1}$ will satisfy the A-truncated τ-property with any even weight function $\tilde{w}$ such that

$$\forall s \geq 0,\ \tilde{w}(s) \leq w\big(\Delta(s)\big),$$

where Δ is defined for any $s \geq 0$ by,

$$\Delta(s) = \inf\big\{|\psi^{-1}(x) - \psi^{-1}(y)|, |x-y| = s,\ y \in A\big\}.$$

We denote by ψ the monotone rearrangement of ν_1 onto ν_α, which is defined by the equation for any $t \in \mathbb{R}$.

$$\nu_1(-\infty, t] = \nu_\alpha(-\infty, \psi(t)].$$

One can easily check that ψ is an odd function, and that its restriction φ on $\mathbb{R}^+$ satisfies,

$$e^{-x} = \int_{\psi(x)}^{+\infty} e^{-u^\alpha} \frac{du}{Z_\alpha},$$

where Z_α is the normalizing constant of μ_α, and sends μ_1 to μ_α. Thus, we are reduced to understand the behavior of the map φ and how it deforms the weights c_ε of Proposition 5.3.3.

5.3.1 Behavior of the monotone rearrangement

When $\alpha \geq 1$, we have the following estimate on the monotone rearrangement due to Talagrand [94].

5.3.6 Lemma ([94, Lemma 2.5])**.** *Let $\alpha \geq 1$. Let ψ be the monotone rearrangement sending ν_1 to ν_α. Denote by Δ the function defined any $s \geq 0$ by,*

$$\Delta(s) = \inf_{|x-y|=s} |\psi^{-1}(x) - \psi^{-1}(y)|. \tag{5.18}$$

There is a constant $c > 0$ depending on α such that for any $s \geq 0$,

$$\Delta(s) \geq c \max(s, s^\alpha).$$

5.3.7 Remark. In [94, Lemma 2.5], this estimate is derived for the monotone rearrangement φ of μ_1 onto μ_α. Since,

$$\forall x \in \mathbb{R},\ \ \psi(x) = \mathrm{sg}(x)\varphi(|x|), \tag{5.19}$$

Lemma 5.3.6 is easily deduced from the same estimate for φ, together with the fact that if x, y have opposite signs,

$$\begin{aligned}\varphi^{-1}(|y|) + \varphi^{-1}(|x|) &\geq c\big(\max(|x|, |x|^\alpha) + \max(|y|, |y|^\alpha)\big)\\ &\geq c' \max(|x-y|, |x-y|^\alpha),\end{aligned}$$

where c' is some constant and where we used the fact that $|x - y| = |x| + |y|$.

To get the exact asymptotic of the tail distribution of ν_α we will need of the following finer estimate on the monotone rearrangement.

5.3.8 Lemma. *Let $\alpha \geq 1$. Define for any $m \geq 1$,*

$$\forall s \geq 0,\ \ \Delta_m(s) = \inf\{|\psi^{-1}(x) - \psi^{-1}(y)| : |x| \leq m,\ \ |x - y| = s\}. \tag{5.20}$$

There is a constant γ depending on α, such that for any $\varepsilon \in (0, 1)$, and $s > m\varepsilon^{-1}$,

$$\Delta_m(s) \geq (1 - \gamma\varepsilon)s^\alpha.$$

Proof. By definition of ψ, we have for any $x \in \mathbb{R}$,

$$\psi^{-1}(x) = -\mathrm{sg}(x) \log \int_{|x|}^{+\infty} e^{-u^\alpha} \frac{du}{Z_\alpha},$$

where Z_α is the normalizing constant of μ_α. Let $s \geq m\varepsilon^{-1}$ and $x, y \in \mathbb{R}$ such that $0 \leq |x| \leq m$, and $|x - y| = s$. If x and y have the same signs, we can assume without loss of generality, that both $x, y \geq 0$. As $x \leq m \leq s$, we have $y = x + s$. Thus,

$$\psi^{-1}(y) - \psi^{-1}(x) \geq \psi^{-1}(s) - \psi^{-1}(m).$$

We have, on one hand, as $s \geq 1$,

$$\int_s^{+\infty} e^{-u^\alpha} du \leq \frac{1}{\alpha} \int_s^{+\infty} \alpha u^{\alpha-1} e^{-u^\alpha} du = \frac{1}{\alpha} e^{-s^\alpha}.$$

And on the other hand,

$$\int_m^{+\infty} e^{-u^\alpha} du \geq e^{-(m+1)^\alpha}.$$

Therefore, as $s > m\varepsilon^{-1}$,

$$\psi^{-1}(y) - \psi^{-1}(x) \geq s^\alpha - (m+1)^\alpha + \log\alpha \geq s^\alpha(1 - \gamma\varepsilon^\alpha),$$

for some constant $\gamma > 0$. Now, if x and y have opposite signs, we can assume without loss of generality that $x \leq 0$ and $y \geq 0$. Then, $y \geq s - m$ so that,

$$|\psi^{-1}(y) - \psi^{-1}(x)| = \psi^{-1}(-x) + \psi^{-1}(y) \geq \psi^{-1}(s-m) \geq (s-m)^\alpha + \log\alpha.$$

Thus, we can find some constant γ' such that $|\psi^{-1}(y) - \psi^{-1}(x)| \geq (1 - \gamma'\varepsilon^\alpha)s^\alpha$. $\square$

5.3.9 Remark. The truncation we performed here is made to ensure we get the best constant (that is 1) in the estimate of the large increments of the monotone rearrangement. Indeed, defining Δ as in (5.18) without the truncation, we would get for $s \gg 1$,

$$\Delta(s) \leq \left|\varphi\Big(\frac{s}{2}\Big) - \varphi\Big(\frac{-s}{2}\Big)\right| = 2\varphi\Big(\frac{s}{2}\Big) \simeq 2\Big(\frac{s}{2}\Big)^\alpha = 2^{1-\alpha}s^\alpha,$$

with $2^{1-\alpha} < 1$.

When $\alpha < 1$, we showed in chapter 6.16 in Lemma 2.2.4, some estimates on the monotone rearrangement of ν_1 onto ν_α. As in the case $\alpha \geq 1$, we can refine this estimate to get the following result.

5.3.10 Lemma. *Let $\alpha \in (0,1)$. Let $\varepsilon \in (0,1)$. Define the function Δ_m by,*

$$\forall s \geq 0,\ \Delta_m(s) = \inf\big\{|\psi^{-1}(y) - \psi^{-1}(x)| : |x| \leq m,\ |x - y| = s\big\}.$$

There is some constant $\kappa > 0$,

$$\Delta_m(s) \geq \begin{cases} \gamma^{-1}(m/\varepsilon)^{\alpha-1}s & \text{if } s < \frac{m}{\varepsilon}, \\ \big(1 - \gamma\varepsilon^{\alpha/2}\big)|s|^\alpha & \text{if } s \geq \frac{m}{\varepsilon}. \end{cases}$$

Proof. Since φ and ψ are linked by the the relation (5.19), the same estimate as in Lemma 2.2.4 holds for the Brenier map ψ. Therefore, we have for any $|x| \leq \psi^{-1}(m)$, and $y \in \mathbb{R}$,

$$|\psi(y) - \psi(x)| \leq K \max\big(\psi^{-1}(m)^{\frac{1}{\alpha}-1}|y-x|, |y-x|^{\frac{1}{\alpha}}\big),$$

with $K \geq 1$. Fix $|x| \leq m$, and $y \in \mathbb{R}$. We have

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq K^{-\alpha}\min(|y-x|^\alpha, \psi^{-1}(m)^{1-\frac{1}{\alpha}}|y-x|),$$

But we know from (2.8) that for $m \geq 1$, $\psi^{-1}(m) \geq c_0 m^\alpha$, with some constant $c > 0$. Thus, for $m \geq 1$, there is a constant $\gamma > 0$, which will vary along the proof without changing name, such that

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq \gamma^{-1}\min(|y-x|^\alpha, m^{\alpha-1}|y-x|).$$

We deduce that for $|x-y| \leq m/\varepsilon$,

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq \gamma^{-1}\Big(\frac{m}{\varepsilon}\Big)^{\alpha-1}|y-x|.$$

Let $s = |x-y|$. Assume now $s \geq m/\varepsilon$. Proceeding as in the proof of Lemma 5.3.8 in the case $\alpha \geq 1$, we assume first that $x, y \geq 0$. As $s \geq m \geq x$, we must have $y = x+s$. Then,

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq \psi^{-1}(s) - \psi^{-1}(m).$$

On one hand, as $\alpha < 1$, we have

$$\int_m^{+\infty} e^{-u^\alpha} du = \int_0^{+\infty} e^{-(u+m)^\alpha} du \geq \Big(\int_0^{+\infty} e^{-u^\alpha} du\Big) e^{-m^\alpha} = \frac{1}{C} e^{-m^\alpha},$$

and on the other hand, by (2.4),

$$\int_s^{+\infty} e^{-u^\alpha} du \leq C s^{1-\alpha} e^{-s^\alpha},$$

where C is some constant depending on α large enough. Thus,

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq s^\alpha - m^\alpha - (1-\alpha)\log s - 2\log C.$$

As $\log s \leq (2/\alpha)s^{\alpha/2}$ for $s \geq 1$, we deduce that

$$|\psi^{-1}(y) - \psi^{-1}(w)| \geq s^\alpha(1 - \gamma\varepsilon^{\alpha/2}).$$

If x and y have opposite signs, we can assume $x \leq 0$ and $y \geq 0$, thus $y \geq s-m$ and we get,

$$\begin{aligned}|\psi^{-1}(y) - \psi^{-1}(x)| &\geq \psi^{-1}(y) \geq \psi^{-1}(s-m)\\ &\geq (s-m)^\alpha - (1-\alpha)\log(s-m) - \log C.\end{aligned}$$

As $s \geq m/\varepsilon$, we deduce

$$|\psi^{-1}(y) - \psi^{-1}(x)| \geq s^\alpha(1 - \gamma\varepsilon^{\alpha/2}),$$

which ends the proof of the claim. □

5.3.2 A family of weights for ν_α

Using transport arguments, we will work in this section at obtaining a family of weights for ν_α which capture its exact tail distribution.

5.3.11 Proposition. *Let $\alpha > 0$, $\alpha \neq 1$, and $m \geq 1$. There exist some constants $\kappa, \varepsilon_0 > 0$ depending on α such that for any $\varepsilon \in (0, \varepsilon_0)$, $(\nu_\alpha, w_{\alpha,\varepsilon}^{(m)})$ satisfies the $[-m, m]$-truncated τ-property where,*

$$w_{\alpha,\varepsilon}^{(m)}(t) = \begin{cases}\kappa^{-1} e^{-(\frac{m}{\varepsilon})^{\alpha/2}} t^2 & \text{if } |t| \leq m\varepsilon^{-1},\\ (1 - \kappa\varepsilon^{(\alpha/2)\wedge 1})|t|^\alpha & \text{if } |t| > m\varepsilon^{-1}.\end{cases}$$

Proof. Let $\varepsilon \in (0,1)$ and $m \geq 1$. Let $\delta > 0$ such that

$$\frac{1}{2}\left(\frac{m}{\varepsilon}\right)^{\alpha} = \frac{2}{\delta^2}.$$

With this choice of δ, we will prove that for $s \geq 0$,

$$w_\delta(\Delta_m(s)) \geq w^{(m)}_{\varepsilon,\alpha}(s),$$

with the appropriate constants κ and ε_0, w_δ defined in Corollary 5.3.4, and where Δ_m is as in (5.20). Using the result of Lemma 5.3.5, this will yield the claim.

Let ε be small enough such that w_δ is non-decreasing. This is possible since $\delta^2 \leq 2\varepsilon^\alpha$. Let $s \geq m/\varepsilon$. If ε is small enough, we have by Lemma 5.3.10 or Lemma 5.3.8,

$$\Delta_m(s) \geq \frac{1}{2}\left(\frac{m}{\varepsilon}\right)^{\alpha} = \frac{2}{\delta^2}.$$

If $\alpha > 1$, then by Lemma 5.3.8 we get, as $\delta^2 \leq 4\varepsilon^\alpha$,

$$w_\delta(\Delta_m(s)) \geq (1-2\delta)(1-\gamma\varepsilon)s^\alpha \geq (1-\kappa\varepsilon^{(\alpha/2)\wedge 1})s^\alpha,$$

for some constant κ which will vary along the proof. Similarly, when $\alpha < 1$, we get by Lemma 5.3.10,

$$w_\delta(\Delta_m(s)) \geq (1-2\delta)(1-\gamma\varepsilon^\alpha \log(1/\varepsilon))s^\alpha \geq (1-\kappa\varepsilon^{\alpha/2})s^\alpha.$$

Now let $s \leq m/\varepsilon$. Assume $\alpha \geq 1$. By Lemma 5.3.6 and the fact that w_δ is non-decreasing, we have

$$w_\delta(\Delta_m(s)) \geq w_\delta(cs).$$

Without loss of generality, we can assume $c \leq 1/2$. Then, as $m\varepsilon^{-1} \leq 4\delta^{-2}$, we have $cs \leq 2\delta^{-2}$, so that we get

$$w_\delta(\Delta_m(s)) \geq \frac{c^2\delta e^{-\frac{1}{\delta}}s^2}{8}.$$

Using the fact that $\delta e^{-\frac{1}{\delta}} \geq c_1 e^{-2/\delta}$, for some constant $c_1 > 0$, we get the claim in the case $\alpha > 1$. Assume now $\alpha < 1$. From Lemma 5.3.10 and the fact that w_δ is non-decreasing, we deduce

$$w_\delta(\Delta_m(s)) \geq w_\delta(\gamma^{-1}(m/\varepsilon)^{\alpha-1}s).$$

Without loss of generality, we can assume that $\gamma \geq 2$. As $m\varepsilon^{-1} \leq 4\delta^{-2}$ and $s \leq m/\varepsilon$, we have

$$\gamma^{-1}(m/\varepsilon)^{\alpha-1}s \leq \frac{2}{\delta^2}.$$

Thus,

$$w_\delta(\Delta(s)) \geq \frac{1}{8}\delta e^{-\frac{1}{\delta}}\left(\gamma^{-1}(m/\varepsilon)^{\alpha-1}s\right)^2 \geq \kappa^{-1}\delta^a e^{-1/\delta},$$

with some $a > 0$. But, we can find some constant $c_2 > 0$ such that

$$\delta^a e^{-1/\delta} \geq c_2 e^{-2/\delta},$$

which, recalling that $(m\varepsilon^{-1})^\alpha = 4\delta^{-2}$ gives the claim.

□

We can now give a proof of Proposition 5.3.1.

Proof of Proposition 5.3.1. As the truncated τ-property tensorizes, and $(\nu_\alpha, w_{\alpha,\varepsilon}^{(m)})$ satisfies the $[-m, m]$-truncated τ-property for $\varepsilon \in (0, \varepsilon_0)$, for some $\varepsilon_0 > 0$ and any $m \geq 1$, by Proposition 5.3.11, we deduce by the tensorization property of the τ-property (see Lemma 5.3.2) that $(\nu_\alpha^n, W_{\alpha,\varepsilon}^{(m)})$ satisfies the mB_{ℓ^∞}-truncated τ-property with $W_{\alpha,\varepsilon}^{(m)}$ defined in (5.17).

□

5.4 Large deviations

We will prove in this section Theorem 5.2.1. As sketched in the introduction, the proof will consist in looking, in a first phase, for large deviations inequalities for ν_α^n and lower bounds estimates of the probability of translates.

As a consequence of the truncated τ-property of Proposition 5.3.1, satisfied by ν_α^n and the weight functions $W_{\alpha,\varepsilon}^{(m)}$, we deduce an isoperimetric-type bound for ν_α^n with respect to the metric $|| \, ||_{\ell^\alpha}$ (or $|| \, ||_{\ell^\alpha}^\alpha$ in the case $\alpha < 1$). This estimate will be of paramount importance to derive the upper bound of Theorem 5.2.1.

5.4.1 Proposition. *Let $\alpha > 0$, $\alpha \neq 2$. Let $r > 0$. Let $v(n)$, $t(n)$ be two sequences going to $+\infty$ as n goes to $+\infty$. Let E and F be Borel subsets of $\mathbb{R}^n$ such that*

$$F + t(n)B_{\ell^2} \subset E, \ \liminf_{n\to+\infty} \nu_\alpha^n(F) > 0.$$

For $\alpha \neq 1$, we assume that

$$(\log n)^{\alpha/2} = o(\log \frac{t(n)^2}{v(n)}),$$

whereas for $\alpha = 1$, we suppose $v(n) = o(t(n)^2)$. Then,

$$\limsup_{n\to+\infty} \frac{1}{v(n)} \log \nu_\alpha^n\big(x \notin E + (rv(n))^{1/\alpha} B_{\ell^\alpha}\big) \leq -r.$$

5.4.2 Remark. For $\alpha = 2$, the Gaussian isoperimetric inequality (see [70, Theorem 2.5]) entails the same result without any further assumption on the speed $v(n)$ or the set E than $\liminf_n \nu_2^n(E) > 0$.

Proof. Before going into the proof per say, we need to relate the enlargements by the weights $W_{\alpha,\varepsilon}^{(m)}$, for which we know that $(\nu_\alpha^n, W_{\alpha,\varepsilon}^{(m)})$ satisfies the τ-property, and therefore a deviation inequality of the type (5.16), to the ℓ^α-balls. This is the object of the following lemma.

5.4.3 Lemma. *Let $\alpha > 0$. With the notation of Proposition 5.3.1, for any $r > 0$, $m \geq 1$ and $\varepsilon \in (0, \varepsilon_0)$,*

$$\big\{W_{\alpha,\varepsilon}^{(m)} \leq r\big(1 - \kappa\varepsilon^{(\alpha/2)\wedge 1}\big)\big\} \subset k_m(\varepsilon)\sqrt{r}B_{\ell^2} + r^{1/\alpha}B_{\ell^\alpha},$$

with $k_m(\varepsilon) = \sqrt{\kappa}e^{\frac{1}{2}(\frac{m}{\varepsilon})^{\alpha/2}}$. Moreover, there is a function $l : \mathbb{R}_+ \to \mathbb{R}_+$, such that

$$\{W_{1,\varepsilon} \leq r(1 - 2\varepsilon)\} \subset l(\varepsilon)\sqrt{r}B_{\ell^2} + rB_{\ell^1}.$$

Proof. We will prove only the first statement, the proof for the second one being similar. Let $y \in \mathbb{R}^n$. By cutting the entries of y, we can find $y_1, y_2 \in \mathbb{R}^n$, such that $y = y_1 + y_2$, for any $i \in \{1, ..., n\}$, $y_1(i)y_2(i) = 0$, and

$$|y_1(i)| \leq \frac{m}{\varepsilon}, \quad |y_2(i)| > \frac{m}{\varepsilon}.$$

By the very definition of $W^{(m)}_{\alpha,\varepsilon}$,

$$\kappa^{-1}e^{-(m/\varepsilon)^{\alpha/2}} \sum_{i=1}^{n} |y_1(i)|^2 = W^{(m)}_{\alpha,\varepsilon}(y_1) \leq W^{(m)}_{\alpha,\varepsilon}(y),$$

and

$$(1 - \kappa\varepsilon^{(\alpha/2)\wedge 1})||y_2||^{\alpha}_{\ell^\alpha} = W^{(m)}_{\alpha,\varepsilon}(y_2) \leq W^{(m)}_{\alpha,\varepsilon}(y).$$

Thus, if we let

$$k_m(\varepsilon)^2 = e^{(m/\varepsilon)^{\alpha/2}}\kappa,$$

and if $W^{(m)}_{\alpha,\varepsilon}(y) \leq r(1 - \kappa\varepsilon^{(\alpha/2)\wedge 1})$, then $||y_1||_{\ell^2} \leq k_m(\varepsilon)\sqrt{r}$, and $||y_2||^{\alpha}_{\ell^\alpha} \leq r$. □

With this lemma proven, we can now give the proof of Proposition 5.4.1. We start with the case $\alpha = 1$. As $v(n) = o(t(n)^2)$, for n large enough, we have $l(\varepsilon)\sqrt{rv(n)} \leq t(n)$. Then, by Lemma 5.4.3, we have

$$F + \{W_{1,\varepsilon} \leq r(1-2\varepsilon)v(n)\} \subset F + t(n)B_{\ell^2} + rv(n)B_{\ell^1}.$$

But by assumption, $F + t(n)B_{\ell^2} \subset E$. Thus,

$$F + \{W_{1,\varepsilon} \leq r(1-2\varepsilon)v(n)\} \subset E + rv(n)B_{\ell^1}.$$

We deduce that,

$$\nu_1^n(x \notin E + rv(n)B_{\ell^1}) \leq \nu_1^n\big(x \notin F + \{W_{1,\varepsilon} \leq r(1-2\varepsilon)v(n)\}\big).$$

As $(\nu_1^n, W_{1,\varepsilon})$ satisfies the τ-property, we have the following deviation inequality (see (5.15)),

$$\nu_1^n(x \notin E + rv(n)B_{\ell^1}) \leq \frac{1}{\nu_1^n(F)} e^{-r(1-2\varepsilon)v(n)}.$$

As $\liminf_n \nu_1^n(F) > 0$, we get

$$\limsup_{n\to+\infty} \frac{1}{v(n)} \log \nu_1^n(x \notin E + rv(n)B_{\ell^1}) \leq -r(1-2\varepsilon).$$

Letting ε going to 0, we get the claim. Let now $\alpha \neq 1$. Let $\varepsilon \in (0, \varepsilon_0)$ and set $m = c(\log n)^{1/\alpha}$, with some $c > 0$ which will be chosen later. By Lemma 5.4.3

$$F + \{W^{(m)}_{\alpha,\varepsilon} \leq r(1 - \kappa\varepsilon^{(\alpha/2)\wedge 1})v(n)\} \subset F + k_n(\varepsilon)\sqrt{rv(n)}B_{\ell^2} + (rv(n))^{1/\alpha}B_{\ell^\alpha}).$$

From the assumption that $(\log n)^{\alpha/2} = o(\log \frac{t(n)}{\sqrt{v(n)}})$ we deduce that for n large enough,

$$\frac{t(n)}{\sqrt{v(n)}} \geq e^{(\frac{(\log n)}{\varepsilon})^{\alpha/2}}.$$

In particular for n large enough,

$$\sqrt{\frac{\kappa}{r}} \frac{t(n)}{\sqrt{v(n)}} \geq e^{\frac{1}{2}(\frac{(\log n)}{\varepsilon})^{\alpha/2}}.$$

Put in another way

$$k_m(\varepsilon)\sqrt{rv(n)} \leq t(n).$$

Thus,

$$F + \{W_{\alpha,\varepsilon}^{(m)} \leq r(1-\kappa\varepsilon^{(\alpha/2)\wedge 1})v(n)\} \subset F + t(n)B_{\ell^2} + (rv(n))^{1/\alpha}B_{\ell^\alpha}.$$

As by assumption $F + t(n)B_{\ell^2} \subset E$, we get

$$\nu_\alpha^n(x \notin E + (rv(n))^{1/\alpha}B_{\ell^\alpha})) \leq \nu_\alpha^n(x \notin F + \{W_{\alpha,\varepsilon}^{(m)} \leq r(1-\kappa\varepsilon)v(n)\}).$$

As $(\nu_\alpha^n, W_{\alpha,\varepsilon}^{(m)})$ satisfies the mB_{ℓ^∞}-truncated τ-property by Proposition 5.3.1, we deduce the following deviation inequality (see (5.16)),

$$\nu_\alpha^n\big(x \notin F + \{W_{\alpha,\varepsilon}^{(m)} \leq r(1-\gamma\varepsilon^{(\alpha/2)\wedge 1})v(n)\}\big) \leq \frac{1}{\nu_\alpha^n(F \cap mB_{\ell^\infty})} e^{-r(1-\gamma\varepsilon^{(\alpha/2)\wedge 1})v(n)}.$$

But,

$$\int ||x||_\infty d\nu_\alpha^n(x) = \int ||x||_\infty d\mu_\alpha^n(x).$$

Let $\Phi = \varphi^{\otimes n}$, where φ is the monotone rearrangement map sending μ_1 to μ_α. Then Φ sends μ_1^n to μ_α^n, so that,

$$\int ||x||_\infty d\mu_\alpha^n(x) = \int ||\Phi(x)||_\infty d\mu_1^n(x).$$

From (2.8), we deduce

$$\int ||\Phi(x)||_\infty d\mu_1^n(x) \leq K(1 + \int ||x||_\infty^{1/\alpha} d\mu_1^n(x)),$$

for some constant $K > 0$. But $\int ||x||_\infty^{1/\alpha} d\mu_1^n(x) \leq c_0 \log n$, with $c_0 \geq 1$. Therefore,

$$\int ||x||_\infty d\mu_\alpha^n(x) \leq 2Kc_0 \log n. \tag{5.21}$$

Thus by Markov's inequality,

$$\nu_\alpha^n(x \notin mB_{\ell^\infty}) \leq \frac{2Kc_0}{c},$$

since we chose $m = c(\log n)^{1/\alpha}$. As $\liminf_n \nu_\alpha^n(F) > 0$ by assumption, we deduce that for c large enough,

$$\liminf_{n\to+\infty} \nu_\alpha^n(F \cap mB_{\ell^\infty}) > 0.$$

Therefore,

$$\limsup_{n\to+\infty} \frac{1}{v(n)} \log \nu_\alpha^n(x \notin E + (rv(n))^{1/\alpha}B_{\ell^\alpha}\}) \leq -r(1-\kappa\varepsilon^{(\alpha/2)\wedge 1}),$$

which gives the claim by taking $\varepsilon \to 0$. □

We show in the next proposition that we can lower bound the probability of translates under ν_α^n.

5.4.4 Proposition. *Let $\alpha \in (0, 2]$. Let $v(n)$ be a sequence going to $+\infty$ as n goes to $+\infty$. Fix some $r > 0$. Let E be some Borel subset of $\mathbb{R}^n$ such that*

$$\liminf_{n\to+\infty} \nu_\alpha^n(E) > 0.$$

(i). For any sequence h_n of elements of $\mathbb{R}^n$,

$$\liminf_{n\to+\infty} \frac{1}{v(n)} \log \nu_\alpha^n(E + v(n)^{1/\alpha} h_n) \geq -\limsup_{n\to+\infty} ||h_n||_{\ell^\alpha}^\alpha.$$

(ii). If $\alpha \in (0, 1]$, then for any sequence $h_n \in \mathbb{R}_+^n$,

$$\liminf_{n\to+\infty} \frac{1}{v(n)} \log \mu_\alpha^n(E + v(n)^{1/\alpha} h_n) \geq -\limsup_{n\to+\infty} ||h_n||_{\ell^\alpha}^\alpha.$$

5.4.5 Remark. On can obtain the estimate (ii) for the measures μ_α with an additional assumption on the speed that $n = o(v(n))$ when $\alpha \in (1, 2]$, which is actually very restrictive in the applications we have in mind. This is one of the reasons of the limitation of Theorem 5.2.3 to the case $\alpha \leq 1$, since we do not know how to produce a meaningful lower bound of such translated sets in this case. Similarly, when $\alpha > 2$, one can see, at least for α integer, that the estimate (i) does not hold unless $n = o(v(n))$.

Proof. The proof will essentially follow the lines of [44, Theorem 5.1]. Indeed, in the Gaussian case $\alpha = 2$, this lower bound is derived from the translation formula of the Gaussian measure. The proof for $\alpha < 2$ will consist in mimicking the Gaussian case.

If the lim sup in the right-hand side of (i) is infinite, then the statement is trivial. If it is finite, we take some $\tau > 0$, such that $||h_n||_{\ell^\alpha}^\alpha \leq \tau$, for all $n \in N$. Let for any $h \in \mathbb{R}^n$, $W_\alpha(h) = \sum_{i=1}^n |h_i|^\alpha$. Then, we have,

$$\nu_\alpha^n(E + v(n)^{1/\alpha} h_n) = \frac{1}{Z_n} \int_E e^{-W_\alpha(y + v(n)^{1/\alpha} h)} d\ell_n(y),$$

where ℓ_n denotes the Lebesgue measure on $\mathbb{R}^n$, and Z_n is the normalizing factor. If $\alpha \in (0, 1)$, then for any $s, t \in \mathbb{R}$,

$$|s + t|^\alpha \leq |s|^\alpha + |t|^\alpha.$$

Thus,

$$W_\alpha(y + v(n)^{1/\alpha} h_n) \leq W_\alpha(y) + v(n) W_\alpha(h_n)$$

Therefore,

$$\nu_\alpha^n(E + v(n)^{1/\alpha} h_n) \geq e^{-v(n) W_\alpha(h_n)} \nu_\alpha^n(E),$$

which gives the claim in the case $\alpha \in (0, 1)$. Note that the same argument for μ_α instead of ν_α gives without changes the estimate (ii).

Now, if $\alpha \in [1,2]$, we have for any $s,t \in \mathbb{R}$,

$$|s+t|^\alpha \le |s|^\alpha + \alpha \mathrm{sg}(st)|s|^{\alpha-1}|t| + |t|^\alpha, \tag{5.22}$$

where $\mathrm{sg}(st)$ stands for the sign of st. Thus, for any $y,h \in \mathbb{R}^n$,

$$W_\alpha(y + v(n)^{1/\alpha}h) \le W_\alpha(y) + \alpha v(n)^{1/\alpha}\Phi(y,h) + v(n)W_\alpha(h),$$

where

$$\Phi(y,h) = \sum_{i=1}^n \varphi(y_i,h_i), \tag{5.23}$$

and $\varphi(y,h) = \mathrm{sg}(yh)|y|^{\alpha-1}|h|$. We have,

$$\begin{aligned}\frac{1}{Z_n}\int_{y\in E} e^{-W_\alpha(y+v(n)^{1/\alpha}h_n)}d\ell_n(y) &\ge \frac{e^{-v(n)W_\alpha(h_n)}}{Z_n}\int_E e^{-W_\alpha(y)-\alpha v(n)^{1/\alpha}\Phi(y,h_n)}d\ell_n(y)\\ &= e^{-v(n)W_\alpha(h_n)}\int_E e^{-\alpha v(n)^{1/\alpha}\Phi(x,h_n)}d\nu_\alpha^n(x).\end{aligned}$$

Jensen's inequality yields,

$$\int_E e^{-\alpha v(n)^{1/\alpha}\Phi(x,h_n)}d\nu_\alpha^n(x) \ge \nu_\alpha^n(E)\exp\Big(-\frac{\alpha v(n)^{1/\alpha}}{\nu_\alpha^n(E)}\int_E \Phi(x,h_n)d\nu_\alpha^n(x)\Big).$$

But, by Cauchy-Schwarz inequality,

$$\int_E \Phi(x,h_n)d\nu_\alpha^n(x) \le \nu_\alpha^n(E)^{1/2}\Big(\int \Phi(x,h_n)^2 d\nu_\alpha^n(x)\Big)^{1/2}.$$

But $\int \varphi(x,h)d\nu_\alpha(x) = 0$ for any $h \in \mathbb{R}$, since $\varphi(-x,h) = -\varphi(x,h)$ and ν_α is symmetric. Thus,

$$\int \Phi(x,h_n)^2 d\nu_\alpha^n(x) = \int |t|^{2(\alpha-1)}d\nu_\alpha^n(t)\big(\sum_{i=1}^n |h_n(i)|^2\big).$$

Using the fact that $\alpha < 2$, we get,

$$\Big(\int \Phi(x,h_n)^2 d\nu_\alpha^n(x)\Big)^{\frac{\alpha}{2}} \le c^{\frac{\alpha}{2}}W_\alpha(h_n),$$

where $c > 0$ is some constant. As $W_\alpha(h_n) \le \tau$, we have

$$\int_{x\in E} e^{-\alpha v(n)^{1/\alpha}\Phi(x,h_n)}d\nu_\alpha^n(x) \ge \nu_\alpha^n(E)\exp\Big(-\frac{c^{1/2}\tau\alpha v(n)^{1/\alpha}}{\nu_\alpha^n(E)^{1/2}}\Big).$$

Note that is was actually very important that we did not bound $\mathrm{sg}(xy)$ by 1 in (5.22), so that $\varphi(.,h)$ is of mean 0 under ν_α, and $\int \Phi(x,h_n)^2 d\nu_\alpha^n(x)$ is not too big. When one replaces ν_α by μ_α, this is exactly where one needs to make an assumption on speed to identify the leading term.

By assumption, we know that there is some $\eta > 0$ such that for n large enough, $\nu_\alpha^n(E) > \eta$. Thus, we get for n large enough,

$$\nu_\alpha^n(x \in E + v(n)^{1/\alpha}h_n) \ge \eta\exp\Big(-v(n)W_\alpha(h_n) - 2\Big(\frac{c}{\eta}\Big)^{1/2}\tau\alpha v(n)^{1/\alpha}\Big).$$

Taking the lim inf at the exponential scale $v(n)$, we get the claim.

□

We can now give a proof of Theorem 5.2.1. We will essentially follow the proof of the LDP of Wiener chaoses (see [44, section 5, Theorem 5.1]), replacing the use of the Cameron-Martin formula by Proposition 5.4.4, and the Gaussian isoperimetric inequality with Proposition 5.4.1.

Proof of Theorem 5.2.1. Without loss of generality we can and will assume that $N = \mathbb{N}$. **Property of the rate function:** By assumption (iv), for any $x \in \mathcal{X}$,

$$I_\alpha(x) = \sup_{\delta>0} \limsup_{n\to+\infty} I_{n,\delta}(x).$$

This formulation shows that $I_\alpha(x) < +\infty$ if and only if there is a sequence $h_n \in \mathbb{R}^n$, $n \in N$, such that

$$\lim_{n\to+\infty} F_n(h_n) = x, \quad \limsup_{n\to+\infty} W_\alpha(h_n) = I_\alpha(x).$$

Thus, $I_\alpha(x) \leq \tau$, for some fixed $\tau \geq 0$, if and only if x is a limit point of a sequence $(F_n(h_n))_{n\in\mathbb{N}}$ such that $\limsup_n W_\alpha(h_n) \leq \tau$. Therefore, I_α is lower semi-continuous. Moreover,

$$\{I_\alpha \leq \tau\} \subset \overline{\cup_{n\in\mathbb{N}} F_n(2\tau B_{\ell^\alpha})}.$$

As by assumption (iv) the set on the right-hand side is compact, we conclude that I_α is a good rate function.

Lower bound: Let $x \in \mathcal{X}$ such that $I_\alpha(x) < +\infty$. By assumption (iv), there is a sequence $h_n \in \mathbb{R}^n$ such that

$$\lim_{n\to+\infty} F_n(h_n) = x, \quad \limsup_{n\to+\infty} W_\alpha(h_n) = I_\alpha(x).$$

Let $\delta > 0$. For n large enough,

$$\mathbb{P}\big(f_n(X_n) \in B(x, 2\delta)\big) \geq \mathbb{P}\big(f_n(X_n) \in B(F_n(h_n), \delta)\big).$$

Let

$$E = \big\{Y \in \mathbb{R}^n : d(f_n(Y + v(n)^{1/\alpha} h_n), F_n(h_n)) < \delta\big\}.$$

Note that

$$\mathbb{P}\big(f_n(X_n) \in B(F_n(h_n), \delta)\big) = \mathbb{P}\big(X_n \in E + v(n)^{1/\alpha} h_n\big).$$

By assumption (i), $\mathbb{P}(X_n \in E)$ goes to 1 as n goes to $+\infty$. From Lemma 5.4.4, we deduce

$$\liminf_{n\to+\infty} \frac{1}{v(n)} \log \mathbb{P}\big(f_n(X_n) \in B(x, 2\delta)\big) \geq -I_\alpha(x).$$

Upper bound: Let A be a closed subset of $\mathcal{X}$. We can assume without loss of generality that $\inf_A I_\alpha > 0$. Let $r > 0$ such that $\inf_A I_\alpha > r$. As I_α is a good rate function, we can find a $\delta > 0$ such that

$$A \cap V_\delta(\{I_\alpha \leq r\}) = \emptyset,$$

where V_δ denotes the δ-neighborhood for the distance d. Thus,

$$\mathbb{P}\big(f_n(X_n) \in A\big) \leq \mathbb{P}\big(f_n(X_n) \notin V_\delta(\{I_\alpha \leq r\})\big).$$

Let

$$V = \big\{ Y \in \mathbb{R}^n : f_n(Y) \in V_\delta(\{I_\alpha \le r\}) \big\}.$$

Define, similarly as for the lower bound, the event

$$E_\delta = \big\{ Y \in \mathbb{R}^n : \sup_{h \in r^{1/\alpha} B_{\ell^\alpha}} d\big(f_n(Y + v(n)^{1/\alpha} h_n), F_n(h_n)\big) < \delta \big\}.$$

By assumption (i), we know that $\mathbb{P}(X_n \in E_\delta)$ goes to 1 as n goes to $+\infty$. Observe that

$$E_\delta + (v(n)r)^{\frac{1}{\alpha}} B_{\ell^\alpha} \subset V.$$

Indeed, if $h_n \in r^{1/\alpha} B_{\ell^\alpha}$ and $Y \in E_\delta$, then $I_\alpha(F_n(h_n)) \le r$, from the definition (5.2) of I_α, and

$$d(f_n(Y + v(n)^{1/\alpha} h_n), F_n(h_n)) < \delta,$$

so that $Y + v(n)^{1/\alpha} h_n \in V$. With this observation we get,

$$\mathbb{P}\big(f_n(X_n) \in A\big) \le \mathbb{P}\big(X_n \notin E_\delta + (v(n)r)^{\frac{1}{\alpha}} B_{\ell^\alpha}\big).$$

If $\alpha = 2$, we get by the Gaussian isoperimetric inequality (see [70, Theorem 2.5]) for any n large enough so that $\mathbb{P}(X_n \in E_\delta) \ge 1/2$,

$$\mathbb{P}\big(X_n \notin E_\delta + (v(n)r)^{\frac{1}{2}} B_{\ell^2}\big) \le e^{-v(n)r},$$

which gives the upper bound. Let now $\alpha < 2$, and $t = t_{\delta/4}$, where $t_{\delta/4}$ is given by assumption (ii). With the notation of Theorem 5.2.1 define,

$$F = E_{\delta/2} \cap \big\{ y \in \mathbb{R}^n : \sup_{||h_n||_{\ell^2} \le t} \mathcal{L}_n(h_n) \le \frac{\delta}{2} \big\}.$$

From the assumptions (i), (ii) and Markov's inequality, we deduce that $\liminf_n \mathbb{P}(X_n \in F) > 0$. Furthermore, we claim that

$$F + tB_{\ell^2} \subset E_\delta. \tag{5.24}$$

Indeed, if $y \in F$ and $k \in tB_{\ell^2}$, then by definition of $\mathcal{L}_n$, for all $h \in rB_{\ell^\alpha}$

$$d\big(f_n(y + v(n)^{1/\alpha} h + k), f_n(y + v(n)^{1/\alpha} h)\big) \le \frac{\delta}{2},$$

which yields (5.24) by triangular inequality. By assumption (ii) the requirements of Lemma 5.4.1 are met, thus

$$\limsup_{n \to +\infty} \frac{1}{v(n)} \log \mathbb{P}\big(f_n(X_n) \in A\big) \le -r.$$

As this inequality is true for any $r < \inf_A I_\alpha$, we get the upper bound.

□

We will end this section with the proof of Theorem 5.2.3.

Proof of Theorem 5.2.3. We will follow the same steps as for the proof of Theorem 5.2.1. The compactness assumption (iii), and the assumption (iv) yields that I_α is a good rate function. As shown in the proof of Theorem 5.2.1, a deviation upper bound holds with speed $v(n)$ and rate function I_α, under the assumptions $(i)-(ii)-(iii)$. Thus, we only have to prove the lower bound. Let $x \in \mathcal{X}$ such that $I_\alpha(x) < +\infty$. Because of assumption $(iv)'$, we know that there is a sequence $h_n \in \mathbb{R}_+^n$ such that

$$\lim_{n\to+\infty} F_n(h_n) = x, \quad \limsup_{n\to+\infty} W_\alpha(h_n) = I_\alpha(x).$$

Proceeding as in the proof of Theorem 5.2.1, if $\delta > 0$, then for n large enough,

$$\mathbb{P}\big(f_n(X_n) \in B(x, 2\delta)\big) \geq \mathbb{P}\big(f_n(X_n) \in B(F_n(h_n), \delta)\big).$$

Let

$$E = \big\{Y \in \mathbb{R}^n : d(f_n(Y + v(n)^{1/\alpha}h_n), F_n(h_n)) < \delta\big\}.$$

Note that

$$\mathbb{P}\big(f_n(X_n) \in B(F_n(h_n), \delta)\big) = \mathbb{P}\big(X_n \in E + v(n)^{1/\alpha}h_n\big).$$

By assumption (i), $\mathbb{P}(X_n \in E)$ goes to 1 as n goes to $+\infty$. From Lemma 5.4.4, we deduce

$$\liminf_{n\to+\infty} \frac{1}{v(n)} \log \mathbb{P}\big(f_n(X_n) \in B(x, 2\delta)\big) \geq -I_\alpha(x),$$

which ends the proof of the theorem. □

5.5 Deterministic equivalents for Wigner matrices

We will prove in this section some uniform deterministic equivalents for the spectral measure and largest eigenvalue of deformed Wigner matrices having concentration $\mathcal{C}_\alpha$ for $\alpha \in (0, 2)$, using the inequalities proved in the preceding section. We will also prove a deterministic equivalent for traces of polynomials of deformed Wigner matrices, but which will not rely on concentration arguments. In particular, these deterministic equivalents will entail that assumption (i) of Theorem 5.2.1 holds for the spectral measure, the largest eigenvalue and the traces of polynomials of Wigner matrices in $\mathcal{S}_\alpha$. More precisely, we will prove the following propositions.

5.5.1 Proposition. *Let $\alpha \in (0, 2)$. Let X be a Wigner matrix such that $\mathbb{E}|X_{1,2} - \mathbb{E}X_{1,2}|^2 = 1$ and satisfying the concentration property $\mathcal{C}_\alpha$. For any $r > 0$,*

$$\sup_{H \in rn^{1/\alpha}B_{\ell^\alpha}} d\big(\mu_{X/\sqrt{n}+H}, \mu_{sc} \boxplus \mu_H\big) \underset{n\to+\infty}{\longrightarrow} 0,$$

in probability.

5.5.2 Remark. This statement fails when $\alpha = 2$ since $X/\sqrt{n}$ is in $rn^{1/2}B_{\ell^2}$ for some $r > 0$, with positive probability uniform in n. Whereas on one hand, by Wigner's theorem (see [3])

$$\mu_{2X/\sqrt{n}} \underset{n\to+\infty}{\rightsquigarrow} \mu_{sc,2},$$

in probability, where for any $a>0$,

$$\mu_{sc,a} = \frac{1}{2a^2\pi}\sqrt{4a^2 - x^2}\mathbb{1}_{|x|\leq 2a}dx.$$

On the other hand, by continuity of the free convolution (see [21, Proposition 4.13]),

$$\mu_{sc} \boxplus \mu_{X/\sqrt{n}} \underset{n\to+\infty}{\rightsquigarrow} \mu_{sc} \boxplus \mu_{sc},$$

in probability, and we have $\mu_{sc} \boxplus \mu_{sc} = \mu_{sc,\sqrt{2}}$ by [3, Example 5.3.26].

5.5.3 Proposition. *Let $\alpha \in (0,2)$. Let X be a Wigner matrix with entries symmetric in law such that $\mathbb{E}|X_{1,2}|^2 = 1$. We assume that X satisfies the concentration property $\mathcal{C}_\alpha$. Define the function ρ by,*

$$\forall x \in \mathbb{R},\ \rho(x) = \begin{cases} x + \frac{1}{x} & \text{if } x \geq 1, \\ 2 & \text{otherwise.} \end{cases} \tag{5.25}$$

For any $r>0$,

$$\sup_{A \in rB_{\ell^\alpha}} \left|\lambda_{X/\sqrt{n}+A} - \rho(\lambda_A)\right| \underset{n\to+\infty}{\longrightarrow} 0,$$

in probability.

For the traces of polynomials of independent Wigner matrices we will prove the next proposition.

5.5.4 Proposition. *Let $\alpha \in (0,2]$. Let $P \in \mathbb{C}\langle \boldsymbol{X}\rangle$ be a non-commutative polynomial of total degree $d > \alpha$. Let $\boldsymbol{X} = (X_1, ..., X_p)$ be a family of independent Wigner matrices. For any $r>0$,*

$$\sup_{\boldsymbol{H}\in rB_{\ell^\alpha}} \left|\tau_n[P(\boldsymbol{X}/\sqrt{n} + n^{1/d}\boldsymbol{H})] - \tau[P(\boldsymbol{s})] - \mathrm{tr}[P_d(\boldsymbol{H})]\right| \underset{n\to+\infty}{\longrightarrow} 0,$$

in probability, where P_d is the homogeneous part of degree d of P, $\boldsymbol{s} = (s_1, ..., s_p)$ is a free family of p semi-circular variables in a non-commutative probability space $(\mathcal{A}, \tau)$ and,

$$B_{\ell^\alpha} = \left\{\boldsymbol{H} \in (\mathcal{H}_n^{(\beta)})^p : \sum_{i=1}^p \mathrm{tr}|H_i|^\alpha \leq 1\right\}.$$

It is interesting to note that we are able for polynomials, to make the approximation hold uniformly in $H \in rB_{\ell^2}$, which is why we can consider the Gaussian case in our large deviations principle of Theorem 5.2.9.

5.5.1 Deterministic equivalents in expectation

Our approach to prove Propositions 5.5.1 and 5.5.3 consists is showing in a first step the proposed uniform deterministic equivalents in expectation, and then make use the concentration inequalities of the last section 2 together with a chaining argument to show that these equivalent hold uniformly in probability.

For the empirical spectral measure, we have such a uniform deterministic equivalents in expectation by the following result of Bordenave and Caputo [29].

5.5.5 Theorem ([29, Theorem 2.6])**.** *Let X be a Wigner matrix such that $\mathbb{E}|X_{1,2} - \mathbb{E}X_{1,2}|^2 = 1$, $\mathbb{E}|X_{1,2}|^3 < +\infty$, and $\mathbb{E}X_{1,1}^2 < +\infty$. There exists a universal constant $c > 0$ such that for any $H \in \mathcal{H}_n^{(\beta)}$,*

$$\delta(\mathbb{E}\mu_{X/\sqrt{n}+H}, \mu_{sc} \boxplus \mu_H) \leq c\frac{\sqrt{\mathbb{E}X_{1,1}^2} + \mathbb{E}|X_{1,2}|^3}{\sqrt{n}},$$

where δ is defined for any $\mu, \nu \in \mathcal{P}(\mathbb{R})$,

$$\delta(\mu, \nu) = \sup\{|g_\mu(z) - g_\nu(z)| : \Im z \geq 1\},$$

where g_μ and g_ν denote the Stieltjes transforms of μ and ν.

For the largest eigenvalue, we will prove the following proposition.

5.5.6 Proposition. *Let $\alpha \in (0, 2]$. Let X be a Wigner matrix with entries symmetric in law such that $\mathbb{E}|X_{1,2}|^2 = 1$. We assume that X satisfies the concentration property $\mathcal{C}_\alpha$. For any $r > 0$,*

$$\sup_{H \in rB_{\ell^2}} |\mathbb{E}\lambda_{X/\sqrt{n}+H} - \rho(\lambda_H)| \underset{n \to +\infty}{\longrightarrow} 0,$$

where ρ is the function defined in (5.25).

5.5.7 Remark. The restriction to Wigner matrices with symmetric entries is by no mean optimal, but it allows us to use the estimate of [35] of the Stieltjes transform of deformed Wigner matrices with this symmetry assumption.

We will divide the proof into two lemmas. First we will compare $\mathbb{E}\lambda_{X/\sqrt{n}+H}$ with the right edge of the support of $\mu_{sc} \boxplus \mu_H$.

5.5.8 Lemma. *Under the assumptions of Proposition 5.5.6, for any $r > 0$,*

$$\sup_{||A||_{\ell^2} \leq r} |\mathbb{E}\lambda_{X/\sqrt{n}+A} - b_{s+A}| \underset{n \to +\infty}{\longrightarrow} 0,$$

where b_{s+A} denotes the supremum of the support of $\mu_{sc} \boxplus \mu_A$.

Proof. In a first step, we will focus on showing that $\mathbb{E}\lambda_{X/\sqrt{n}}$ cannot be much larger than b_{s+A} uniformly in $||A|| \leq r$. This will be done by using the estimate of the Stieltjes transform of deformed Wigner matrices proved by Capitaine, Donati-Martin, Féral and Février in [35], and by following a rather classical argument relying on the Helffer-Sjöstrand formula. Our main task will be to make sure that we get a uniform convergence. In a second part, we will get the other side of the convergence, meaning that $\mathbb{E}\lambda_{X/\sqrt{n}}$ cannot be much smaller than b_{s+A} uniformly in $A \in rB_{\ell^2}$. This will fall from the fact that the empirical spectral measure of $X/\sqrt{n} + A$ is asymptotically close to $\mu_{sc} \boxplus \mu_A$ uniformly in $A \in rB_{\ell^2}$, and from the lower semi-continuity of the "right-edge of the support map".

Let $A \in \mathcal{H}_n^{(\beta)}$ and denote by g_A the Stieltjes transform of its spectral measure. We know from [23, Corollary 3] that $b_{s+A} = \mathcal{H}_A(u_A)$, where

$$\forall z \notin \mathrm{supp}(\mu_A),\ \mathcal{H}_A(z) = z + g_A(z), \tag{5.26}$$

and

$$u_A = \sup\{t \in \mathbb{R} : -g'_A(t) \geq 1\}. \tag{5.27}$$

Note that since μ_A is an atomic measure, $-g'_A(u_A) = 1$. We start with a truncation argument to restrain our study of the deviations of $\lambda_{X/\sqrt{n}+A}$ in a compact subset. Denote by W the deformed matrix $X/\sqrt{n} + A$. Note that $-g'_A(\lambda_A) = +\infty$ and $-g'_A(\lambda_A + 1) \leq 1$. As $-g'_A$ is non-increasing, we deduce

$$\lambda_A < u_A \leq \lambda_A + 1. \tag{5.28}$$

As $\mathcal{H}_A$ is non-decreasing on $(u_A, +\infty)$, by definition of u_A, we have

$$\lambda_A < u_A \leq b_{s+A} \leq \lambda_A + 2. \tag{5.29}$$

We deduce that b_{s+A} is bounded uniformly for $||A|| \leq r$. Besides, $|\lambda_W| \leq ||W|| \leq r + ||X/\sqrt{n}||$. Note that with the same argument as for the largest eigenvalue, one can deduce a concentration inequality for $||X/\sqrt{n}||$ similar to the one of Propositions 2.5.1. In particular, it entails that $||X/\sqrt{n}||$is uniformly integrable. Therefore, it is sufficient to show that for any fixed $C > 0$,

$$\limsup_{n\to+\infty} \sup_{||A||\leq r} |\mathbb{E}\lambda_W - b_{s+A}|\mathbb{1}_{|\lambda_W|\leq C} = 0. \tag{5.30}$$

We will begin by the "right-limit" of $(\lambda_W - b_{s+A})_+\mathbb{1}_{|\lambda_W|\leq C}$. Let g_n and g denote respectively the Stieltjes transforms of μ_W and $\mu_{sc} \boxplus \mu_A$. We know from [35, Proposition 4.1, Lemma 5.1] that if X has entries which are symmetric in law and which satisfy a Poincaré inequality, then we have for any $z \in \mathbb{C}^+$,

$$\mathbb{E}g_n(z) = g(z) + \frac{E(z)}{n} + \delta_n(z), \tag{5.31}$$

where $\delta_n(z) = O(n^{-2}|\Im z|^{-d})$, for some $d \geq 1$, can be taken uniform in $A \in \mathcal{H}_n^{(\beta)}$ such that $||A|| \leq r$, and where E is, for n large enough, the Stieltjes transform of a compactly supported measure Λ whose support is included in the one of $\mu_{sc} \boxplus \mu_A$.

One can see, from the concentration inequality which Lipschitz functions of X verify by Lemmas 2.2.1 or 2.2.6, that the same bounds on the variance of variables involving coefficients of the resolvent as in [35, Lemma 9.1] holds, possibly with a $(\log n)^\zeta$ factor, with $\zeta > 0$, in the case where $\alpha < 1$. Therefore, under our assumptions, we have the estimate (5.31) with the error term,

$$\delta_n = O(n^{-2}(\log n)^\zeta |\Im z|^{-d}),$$

for some $d \geq 1$, which can be taken uniform in $A \in \mathcal{H}_n^{(\beta)}$ such that $||A|| \leq r$.

Let $C > 2 + r$ and $\varepsilon > 0$. Let θ be some non-negative function supported in $[b_{s+A}, C]$ of class C^{d+1}, and such that

$$\forall x \in [b_{s+A} + \varepsilon, 2 + r],\ \theta(x) = x - b_{s+A}.$$

We know by the Helffer-Sjöstrand formula [3, (5.5.11)], that for any $\mu \in \mathcal{P}(\mathbb{R})$,

$$\int \theta d\mu = \int_{\mathbb{C}^+} \Re\big(\overline{\partial}\Theta(z) g_\mu(z)\big) dz, \tag{5.32}$$

with $\Theta(x+iy) = \sum_{j=0}^{d+1} \frac{(iy)^j}{j!}\theta^{(j+1)}(x)\chi(y)$ for any $x, y \in \mathbb{R}$, where χ is a smooth compactly supported function which is equal to 1 in a neighborhood of 0. For y in this neighborhood, we have for any $x \in \mathbb{R}$,

$$\overline{\partial}\Theta(x+iy) = \frac{(iy)^d}{d!}\theta^{(d+1)}(x).$$

This shows that the integral on the right-hand side of (5.32) is well defined. As θ is supported on the complement of the support of $\mu_{sc} \boxplus \mu_A$, we have, using (5.32),

$$\mathbb{E}\int \theta d\mu_W = \int_{\mathbb{C}^+} \Re\big[\overline{\partial}\Theta(z)\big(\mathbb{E}g_n(z) - g(z)\big)\big]dz.$$

As E is the Stieltjes transform of a measure whose support is included in $(-\infty, b_{s+A}]$, we get, using again (5.32),

$$\int_{\mathbb{C}^+} \Re\big(\overline{\partial}\Theta(z)E(z)\big)dz = 0.$$

But for $\Im z$ in a neighborhood of 0, $\overline{\partial}\Phi(z)\Theta_n(z) = O(n^{-2}(\log n)^{\zeta})$, therefore as Θ is compactly supported, we deduce that

$$\mathbb{E}\int \theta d\mu_W = \int_{\mathbb{C}^+} \Re\big[\overline{\partial}\Theta(z)\delta_n(z)\big]dz = O\Big(\frac{(\log n)^{\zeta}}{n^2}\Big).$$

In particular,

$$\mathbb{E}\big((\lambda_W - b_{s+A})_+ \mathbb{1}_{\lambda_W \in [b_{s+A}+\varepsilon, C]}\big) \leq \mathbb{E}\theta(\lambda_W) \leq n\int \theta d\mu = O\Big(\frac{(\log n)^{\zeta}}{n}\Big).$$

Thus,

$$\begin{aligned}\mathbb{E}\big(\lambda_W - b_{s+A})_+ \mathbb{1}_{\lambda_W \leq C}\big) &\leq \varepsilon + \mathbb{E}\big((\lambda_W - b_{s+A})_+ \mathbb{1}_{\lambda_W \in [b_{s+A}+\varepsilon, C]}\big) \\ &\leq \varepsilon + O\Big(\frac{(\log n)^{\zeta}}{n}\Big).\end{aligned}$$

Taking the limsup as n goes to $+\infty$ and then ε to 0, we get

$$\lim_{n\to+\infty} \sup_{||A||\leq r} \mathbb{E}(\lambda_W - b_{s+A})_+ \mathbb{1}_{\lambda_W \leq C} = 0.$$

Therefore, by Jensen's inequality

$$\lim_{n\to+\infty} \sup_{||A||\leq r} \big(\mathbb{E}(\lambda_W - b_{s+A})\mathbb{1}_{|\lambda_W| \leq C}\big)_+ = 0. \tag{5.33}$$

It remains to show that

$$\lim_{n\to+\infty} \inf_{||A||_{\ell^2}\leq r} \mathbb{E}(\lambda_W - b_{s+A})\mathbb{1}_{|\lambda_W| \leq C} \geq 0.$$

By Fatou's lemma, it is enough to prove,

$$a.s \ \liminf_{n\to+\infty} \inf_{||A||_{\ell^2}\leq r} \big(\lambda_W - b_{s+A}\big) \geq 0. \tag{5.34}$$

As $\mu_{X/\sqrt{n}}$ converges almost surely to μ_{sc} by Wigner's theorem, we deduce by Hoeffman-Wielandt inequality (see Lemma 2.4.1) that for fixed $r > 0$,

$$a.s \sup_{||A||_{\ell^2} \leq r} d(\mu_W, \mu_{sc} \boxplus \mu_A) \underset{n \to +\infty}{\longrightarrow} 0. \tag{5.35}$$

Note that the map

$$\mu \in \mathcal{P}(\mathbb{R}) \mapsto \sup \mathrm{supp}(\mu) := b_\mu \in \mathbb{R} \cup \{+\infty\},$$

is lower semicontinuous with respect to the weak topology. Besides, as μ_{sc} has zero mean, $\mu_{sc} \boxplus \mu_A(x^2) = \mu_{sc}(x^2) + \mu_A(x^2)$, and therefore

$$\{\mu_{sc} \boxplus \mu_A : ||A||_{\ell^2} \leq r\} \subset \{\mu \in \mathcal{P}(\mathbb{R}) : \mu(x^2) \leq r + 1\}.$$

By Prokhorov's theorem, we deduce that the set on the left-hand side above is pre-compact. Therefore the lower semicontinuity of $\mu \mapsto b_\mu$ can be made uniform on the subset of measures we are looking at, which yields (5.34). □

Finally, we prove that we can approximate b_{s+A} by $\rho(\lambda_A)$ uniformly in A with bounded Hilbert-Schmidt norm, which will give us, together with the preceding lemma, the uniform equivalent of Proposition 5.5.6.

5.5.9 Lemma. *For any $r > 0$,*

$$\sup_{A \in rB_{\ell^2}} |b_{s+A} - \rho(\lambda_A)| \underset{n \to +\infty}{\longrightarrow} 0,$$

where b_{s+A} denotes the supremum of the support of $\mu_{sc} \boxplus \mu_A$, and ρ is as in (5.25).

Proof. Note that we have,

$$\lim_{n \to +\infty} \sup_{A \in rB_{\ell^2}} \mathcal{W}_2(\mu_A, \delta_0) = 0, \tag{5.36}$$

The lower semi-continuity of the map $\mu \mapsto \sup \mathrm{supp}\mu$ for the weak topology gives,

$$\liminf_{n \to +\infty} \inf_{A \in rB_{\ell^2}} b_{s+A} \geq 2, \tag{5.37}$$

as $\mathrm{supp}(\mu_{sc}) = [-2, 2]$. As we saw in the proof of Lemma 5.5.8, $b_{s+A} = \mathcal{H}_A(u_A)$, where u_A and $\mathcal{H}_A$ are defined in (5.27) and (5.26). Moreover $u_A > \lambda_A$. Thus,

$$g_A(u_A) = \int \frac{d\mu_A(x)}{(u_A - x)_+} = \int \frac{1}{x_-} d\mu_A * \delta_{-u_A}(x),$$

where x_- denotes the negative parts of x. Observe that the family $(\delta_{-u_A})_{A \in rB_{\ell^2}}$ is pre-compact since u_A is bounded as $A \in rB_{\ell^2}$, by (5.28). Furthermore, note that for any $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and $u \in \mathbb{R}$,

$$\mathcal{W}_2(\mu * \delta_u, \nu * \delta_u) = \mathcal{W}_2(\mu, \nu).$$

From (5.36), we can deduce in particular that

$$\lim_{n \to +\infty} \sup_{A \in rB_{\ell^2}} d(\mu_A * \delta_{-u_A}, \delta_{-u_A}) = 0, \tag{5.38}$$

using the comparison (2.13) between d and $\mathcal{W}_2$. But the map $\mu \in \mathcal{P}(\mathbb{R}) \mapsto \int t_-^{-1} d\mu(t)$, taking values in $[0, +\infty]$, is lower semi-continuous with respect to the weak topology. By (5.38) and the pre-compactness of $(\delta_{-u_A})_{A \in rB_{\ell^2}}$, we deduce that for any $\delta > 0$,

$$\liminf_{n \to +\infty} \inf_{A \in rB_{\ell^2}} \Big(g_A(u_A) - \frac{1}{\delta \vee (u_A)^+} \Big) \geq 0.$$

But $b_{s+A} = u_A + g_A(u_A)$, therefore

$$\liminf_{n \to +\infty} \inf_{A \in rB_{\ell^2}} \Big(b_{s+A} - u_A - \frac{1}{\delta \vee u_A} \Big) \geq 0.$$

As $u_A > \lambda_A$ and ρ is non-decreasing, we deduce together with (5.37) that

$$\liminf_{n \to +\infty} \inf_{A \in rB_{\ell^2}} \big(b_{s+A} - \rho(\lambda_A) \big) \geq 0.$$

We will prove now that

$$\limsup_{n \to +\infty} \sup_{A \in rB_{\ell^2}} \big(b_{s+A} - \rho(\lambda_A) \big) \leq 0.$$

Let $\varepsilon > 0$ and define

$$E = \{ z \in \mathbb{C} : \Re z \geq \max(\lambda_A, 1) + \varepsilon \}.$$

Then, for any $A \in \mathcal{H}_n^{(\beta)}$ and $z \in E$,

$$|g_A(z) - g_0(z)| \leq \int \Big| \frac{t}{(z-t)z} \Big| d\mu_A(t) \leq \frac{1}{\varepsilon} \int |t| d\mu_A(t) \leq \frac{1}{\varepsilon} \Big(\int t^2 d\mu_A(t) \Big)^{1/2}.$$

Similarly, we have for any $A \in \mathcal{H}_n^{(\beta)}$ and $z \in E$,

$$|g_A'(z) - g_0'(z)| \leq \int \frac{|-t^2 + 2tz|}{|z^2(t-z)^2|} d\mu_A(t) \leq \frac{3}{\varepsilon^2} \int |t| d\mu_A(t) \leq \Big(\int t^2 d\mu_A(t) \Big)^{1/2},$$

as $|t| \leq \Re z \leq |z|$ for any $t \in \operatorname{supp}(\mu_A)$. Therefore, for any $A \in rB_{\ell^2}$, We deduce that for $0 < \varepsilon < 1$,

$$\max \big(||g_A - g_0||_E, ||g_A' - g_0'||_E \big) \leq \frac{3r}{\varepsilon^2 \sqrt{n}}, \tag{5.39}$$

where $||\ ||_E$ denotes the sup norm on E. Let $v_A = \max(\lambda_A, 1)$. As $-g_0'(v_A + \varepsilon) < 1$, we deduce by (5.39) that for n large enough, $-g_A'(v_A + \varepsilon) < 1$, uniformly in $A \in rB_{\ell^2}$. As $-g_A'$ is non-increasing, $u_A \leq v_A + \varepsilon$. But $\mathcal{H}_A$ is increasing on $[u_A, +\infty)$, therefore

$$b_{s+A} \leq \mathcal{H}_A(v_A + \varepsilon) \leq \rho(\lambda_A) + \varepsilon + \frac{3r}{\varepsilon^2 \sqrt{n}},$$

which concludes the proof. □

5.5.2 A chaining argument

We will now give a proof of Propositions 5.5.1 and 5.5.3. As it will rely on a chaining argument, we will need the following lemma.

5.5.10 Lemma. *Let $m \in \mathbb{N}$ and let B_{ℓ^p} denote the ℓ^p-ball of $\mathbb{C}^m$ for any $p > 0$. Fix some $0 < p < q < \infty$. We denote by $N(B_{\ell^p}, \varepsilon B_{\ell^q})$, the covering number of B_{ℓ^p} by εB_{ℓ^q}, that is, the minimal number of translates of εB_{ℓ^q} needed to cover B_{ℓ^p}. There is a constant $c > 0$ depending on p, q, such that for $c(\frac{\log m}{m})^{\frac{1}{p}-\frac{1}{q}} \leq \varepsilon \leq c^{-1}$,*

$$\log N(B_{\ell^p}, \varepsilon B_{\ell^q}) \leq c\varepsilon^{\frac{1}{q}-\frac{1}{p}} \log m.$$

Proof. This estimate is a consequence of the upper bound on entropy numbers of embeddings of ℓ_p^m in ℓ_q^m given in [48, Proposition 3.2.2]. Let $0 < p < q < \infty$. Denote by ℓ_p^m the space $\mathbb{R}^m$ equipped with the (quasi)-norm $|| \, ||_{\ell_p}$. We define, for $k \in \mathbb{N}$,

$$e_k(\ell_p^m \to \ell_q^m) = \inf\{\varepsilon > 0 : B_{\ell_p} \text{ can be covered by } 2^{k-1} \text{ balls } \varepsilon B_{\ell_q}\}.$$

From [48, Proposition 3.2.2], we know that there is a constant $c > 0$ such that for $\log_2(2m) \leq k \leq 2m$,

$$e_k(\ell_p^m \to \ell_q^m) \leq c\Big(k^{-1} \log_2\big(1 + \frac{2m}{k}\big)\Big)^{\frac{1}{p}-\frac{1}{q}}.$$

Thus, if we set $k = \lambda \log_2(2m)$, for some $\lambda \geq 1$ such that $k \leq 2m$, we deduce the following rough bound,

$$e_k(\ell_p^m \to \ell_q^m) \leq c'\lambda^{\frac{1}{q}-\frac{1}{p}},$$

for some constant $c' > 0$. Let now $\varepsilon > 0$ and set λ such that $\varepsilon = c'\lambda^{\frac{1}{q}-\frac{1}{p}}$. The above inequality tells us that if $1 \leq \lambda \leq 2m/\log_2(2m)$, then there are $(2m)^\lambda$ balls εB_{ℓ^q} covering B_{ℓ^p}, that is,

$$N(B_{\ell^p}, \varepsilon B_{\ell^q}) \leq (2m)^\lambda,$$

which yields the claim. □

We are now ready to give a proof of Proposition 5.5.1 and 5.5.3.

Proof of Proposition 5.5.1. Let $H \in \mathcal{H}_n^{(\beta)}$. As X satisfies $\mathcal{C}_\alpha$ for some constant $\kappa > 0$, we see that $X + \sqrt{n}H$ also satisfies $\mathcal{C}_\alpha$ with the same constant κ. We know from Propositions 2.3.1 and 5.5.5, that for any $t > 0$,

$$\mathbb{P}\Big(d\big(\mu_{X/\sqrt{n}+H}, \mu_{sc} \boxplus \mu_H\big) > t + \varepsilon_n\Big) \leq \frac{32}{t^2} \exp\big(-c_\alpha h_\alpha(t)\big),$$

with h_α defined in Proposition 2.3.1 and $\varepsilon_n = O\big(n^{-1/2}(\log n)^{(1/\alpha-1)_+}\big)$, uniformly in $H \in \mathcal{H}_n^{(\beta)}$. Note that the map

$$S : H \in \mathcal{H}_n^{(\beta)} \mapsto d\big(\mu_{X/\sqrt{n}+H}, \mu_{sc} \boxplus \mu_H\big),$$

is $n^{-1/2}$-Lipschitz with respect to $|| \, ||_{\ell^2}$ by Lemma 2.4.1. We deduce using an ε-net argument that for n large enough,

$$\mathbb{P}\big(\sup_{H \in rn^{1/\alpha} B_{\ell^\alpha}} S(H) > 2t \big) \leq \frac{32}{t^2} N(rn^{1/\alpha} B_{\ell^\alpha}, tn^{1/2} B_{\ell^2}) e^{-\frac{c_\alpha}{\kappa^\alpha}(t-\varepsilon_n)_+^\alpha n^{1+\alpha/2}}, \quad (5.40)$$

where $N(rn^{1/\alpha} B_{\ell^\alpha}, tn^{1/2} B_{\ell^2})$ denotes the covering number of $rn^{1/\alpha} B_{\ell^\alpha}$ by $tn^{1/2} B_{\ell^2}$. But, the homogeneity of the norm gives,

$$N(rn^{1/\alpha} B_{\ell^\alpha}, tn^{1/2} B_{\ell^2}) = N(B_{\ell^\alpha}, t' n^{\frac{1}{2} - \frac{1}{\alpha}} B_{\ell^2}),$$

with $t' = t/r$. We get from Lemma 5.5.10 applied with $m = n^2$,

$$\log N(B_{\ell^\alpha}, t' n^{\frac{1}{2} - \frac{1}{\alpha}} B_{\ell^2}) = O(n \log n),$$

This shows that the covering number is negligible with respect to the speed of the deviations, which concludes the chaining argument. □

We finally give a proof of Proposition 5.5.3.

Proof of Proposition 5.5.3. Let $r > 0$. Similarly as in the proof of Proposition 5.5.1, we deduce from Propositions 2.5.1 and 5.5.6, that for any $A \in \mathcal{H}_n^{(\beta)}$ and $t > 0$,

$$\mathbb{P}\Big(\big| \lambda_{X/\sqrt{n}+A} - \rho(\lambda_A) \big| > t + \delta_n \Big) \leq 8 \exp\big(- c_\alpha h_\alpha(t) \big),$$

where h_α is defined in Proposition 2.5.1, $\delta_n = O(n^{-1/2} (\log n)^{(1/\alpha - 1)_+})$ uniformly in $A \in rB_{\ell^2}$, and ρ is as in (5.25). In particular, the error term is uniform for $A \in rB_{\ell^\alpha}$.

Note that the map $x \mapsto \rho(x)$ is 1-Lipschitz. From Weyl's inequality (1.6) and (5.28), we deduce that

$$A \mapsto |\lambda_{X/\sqrt{n}+A} - b_{s+A}|,$$

is 2-Lipschitz with respect to the Hilbert-Schmidt norm on $\mathcal{H}_n^{(\beta)}$. Using an ε-net argument as in the proof of Proposition 5.5.1, it is sufficient to prove that for any fixed $t > 0$, the covering number $N(B_{\ell^\alpha}, tB_{\ell^2})$ is negligible at the exponential scale $n^{\alpha/2}$, that is

$$\log N(B_{\ell^\alpha}, tB_{\ell^2}) = o(n^{\alpha/2}).$$

But from Lemma 5.5.10, we know that,

$$\log N(B_{\ell^\alpha}, tB_{\ell^2}) = O(\log n),$$

which ends the proof of the claim. □

5.5.3 Traces of polynomials of deformed Wigner matrices

We will now prove Proposition 5.5.4. Contrary to the spectral measure or the largest eigenvalue, the proof will consist in a simple moment computation.

Proof of Lemma 5.5.4. By linearity it is sufficient to show the statement when P is a monomial, which we will assume from now on. We can write $P = X_{i_1}...X_{i_q}$, with $q \leq d$. Define the matrix Q with coefficients in $\mathbb{C}\langle \mathbf{X} \rangle$, by

$$Q = \begin{pmatrix} 0 & X_{i_1} & & \\ & \ddots & \ddots & \\ & & \ddots & X_{i_{q-1}} \\ X_{i_q} & & & 0 \end{pmatrix}.$$

Observe that by cyclicity of the trace, for any $\mathbf{Y} \in (\mathcal{H}_n^{(\beta)})^p$, $\mathrm{tr} Q(\mathbf{Y})^q = q \mathrm{tr} P(\mathbf{Y})$. Therefore,

$$\mathrm{tr} P(\mathbf{X}/\sqrt{n} + n^{1/d}\mathbf{H}) = \frac{1}{q} \mathrm{tr}\big(Q(\mathbf{X}/\sqrt{n}) + n^{1/d} Q(\mathbf{H})\big)^d. \tag{5.41}$$

Write $Z = Q(\mathbf{X}/\sqrt{n})$ and $K = Q(\mathbf{H})$. We know from the proof of [7, Lemma 2.1] that,

$$\left|\mathrm{tr}\big(Z + n^{1/d}K\big)^q - \mathrm{tr} Z^q - n^{\frac{q}{d}} \mathrm{tr} K^q\right| \leq 2^q \max_{1 \leq k \leq q-1} n^{\frac{q-k}{d}} (\mathrm{tr}|Z|^{q+1})^{\frac{k}{q+1}} (\mathrm{tr}|K|^2)^{\frac{q-k}{2}}.$$

Let us define q-Schatten (quasi-)norm on $(\mathcal{H}_n^{(\beta)})^p$, for any $q > 0$ by,

$$\forall \mathbf{H} \in (\mathcal{H}_n^{(\beta)})^p,\ ||\mathbf{H}||_q = \left(\sum_{i=1}^p \mathrm{tr}|H_i|^q\right)^{1/q}. \tag{5.42}$$

Note that for any $\mathbf{Y} \in (\mathcal{H}_n^{(\beta)})^p$,

$$|Q(\mathbf{Y})| = \begin{pmatrix} |Y_{i_1}| & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & |Y_{i_q}| \end{pmatrix}.$$

Thus, for any $m \in \mathbb{N}$,

$$\mathrm{tr}|Q(\mathbf{Y})|^m = \sum_{j=1}^q \mathrm{tr}|Y_{i_j}|^m \leq \sum_{i=1}^p \mathrm{tr}|Y_i|^m = ||\mathbf{Y}||_m^m.$$

As $\mathbf{H} \in rB_{\ell^2}$, $\mathrm{tr}|K|^2 \leq r^2$. Without loss of generality we can assume $r \geq 1$. Thus,

$$\left|\mathrm{tr}\big(Z + n^{1/d}K\big)^q - \mathrm{tr}(Z^q) - n^{\frac{q}{d}} \mathrm{tr} K^q\right| \leq r^q 2^q \max_{1 \leq k \leq q-1} n^{\frac{q-k}{d}} ||\mathbf{X}/\sqrt{n}||_{q+1}^k. \tag{5.43}$$

But we know from Wigner's theorem (see [3, Lemma 2.1.6]), that there is a constant $c \geq 1$, such that

$$\mathbb{E}||\mathbf{X}/\sqrt{n}||_{q+1}^{q+1} \leq cn.$$

Besides,

$$\mathbb{E} \max_{1\le k\le q-1} n^{-\frac{k}{d}}||\mathbf{X}/\sqrt{n}||_{q+1}^k \le \sum_{k=1}^{q-1} n^{-\frac{k}{q}}\mathbb{E}||\mathbf{X}/\sqrt{n}||_{q+1}^k.$$

By Jensen's inequality, we deduce

$$\mathbb{E} \max_{1\le k\le q-1} n^{-\frac{k}{d}}||\mathbf{X}/\sqrt{n}||_{q+1}^k \le \sum_{k=1}^{q-1} n^{-\frac{k}{q}}\left(\mathbb{E}||\mathbf{X}/\sqrt{n}||_{q+1}^{q+1}\right)^{\frac{k}{q+1}}.$$

Therefore,

$$\mathbb{E} \max_{1\le k\le q-1} n^{-\frac{k}{d}}||\mathbf{X}/\sqrt{n}||_{d+1}^k \le qcn^{-(\frac{1}{q}-\frac{1}{q+1})}.$$

We deduce from (5.41) and (5.43) that

$$\left|\tau_n[P(\mathbf{X}/\sqrt{n}+n^{1/d}\mathbf{H})] - \mathbb{E}\tau_n[P(\mathbf{X}/\sqrt{n})] - n^{\frac{q}{d}-1}\mathrm{tr}[P(\mathbf{H})]\right| \underset{n\to+\infty}{\longrightarrow} 0,$$

uniformly in $\mathbf{H}\in rB_{\ell^2}$ and where $\tau_n = \frac{1}{n}\mathrm{tr}$.

It is now sufficient to prove that $n^{q/d-1}\mathrm{tr}P(\mathbf{H})$ converges to 0 uniformly in $\mathbf{H}\in rB_{\ell^\alpha}$, as soon as $q<d$. Assume first $q\ge\alpha$. Recall that we proved using the non-commutative Hölder's inequality and the arithmetic-geometric mean inequality, in (2.38),

$$\mathrm{tr}[P(\mathbf{H})] \le \frac{1}{q}\sum_{j=1}^{q}\mathrm{tr}|H_{i_j}|^q. \tag{5.44}$$

As $q\ge\alpha$, we deduce

$$\mathrm{tr}[P(\mathbf{H})] \le \frac{1}{q}||\mathbf{H}||_\alpha^q. \tag{5.45}$$

We conclude that when $\alpha\le q<d$,

$$\sup_{\mathbf{H}\in rB_{\ell^\alpha}} n^{\frac{q}{d}-1}\mathrm{tr}[P(\mathbf{H})] \underset{n\to+\infty}{\longrightarrow} 0.$$

If $q<\alpha$, then $q=1$ and $\alpha>1$. By Jensen's inequality,

$$|\mathrm{tr}H_{i_1}| \le n^{1-1/\alpha}(\mathrm{tr}|H_{i_1}|^\alpha)^{1/\alpha}.$$

Thus, as $d>\alpha$,

$$\sup_{\mathbf{H}\in rB_{\ell^\alpha}} n^{\frac{1}{d}-1}\mathrm{tr}[P(\mathbf{H})] \underset{n\to+\infty}{\longrightarrow} 0.$$

Besides, we know by [3, Theorem 5.4.2], that

$$\mathbb{E}\tau_n[P(\mathbf{X}/\sqrt{n})] \underset{n\to+\infty}{\longrightarrow} \tau[P(\mathbf{s})],$$

where $\mathbf{s}$ are a family of p free semi-circular variables defined on a non-commutative probability space $(\mathcal{A},\tau)$. This ends the proof of the lemma. $\square$

5.6 Deterministic equivalent for the last-passage time

We will prove in this section the analogue of the results for Wigner matrices of the preceding section, for the last-passage time. More precisely, we will provide a deterministic equivalent for the last-passage time when the matrix of weights is deformed by some matrix nH, where $||H||_{\ell^\alpha}$ is bounded for some $\alpha \in (0,1)$.

Let $\mathcal{A}$ denote the set of finite vectors $(v_1, ..., v_m)$, which we will call *admissible*, such that $v_i \in \{0, ..., n\}^d$, $v_0 = (0, ..., 0)$, $v_m = (n, ..., n)$, and for any $i \in \{0, ..., m-1\}$, $v_i < v_{i+1}$, where $<$ denotes the lexicographic order. With this definition we set, for any $H \in \mathbb{R}^I$, where $I = \{0, ..., n\}^d$,

$$\mathcal{T}_n(H) = \sup_{V \in \mathcal{A}} \Big\{ \sum_{i=0}^{m} H_{v_i}^+ + \sum_{i=0}^{m-1} g\Big(\frac{v_{i+1} - v_i}{n}\Big) \Big\}, \tag{5.46}$$

where $V = (v_0, ..., v_m)$ for some $m \in \mathbb{N}$, where g is as in (5.10), and where x^+ denotes here the positive part of $x \in \mathbb{R}$. In the sequel, we will extend the definition of the last-passage time to multi-matrices with real coefficients by setting $T(Y)$ to be equal to $T(Y^+)$ where $Y^+ = (Y_v^+)_{v \in \{0,...,n\}^d}$. With this notation, we will prove the following proposition.

5.6.1 Proposition. *Let $\alpha \in (0,1)$. Let $X = (X_v)_{v \in \mathbb{Z}_+^d}$ be a family of i.i.d random variables following the law μ_α. For any $r > 0$,*

$$\sup_{||H||_{\ell^\alpha} \le r} \Big| \frac{1}{n} T(X + nH) - \mathcal{T}_n(H) \Big| \underset{n \to +\infty}{\longrightarrow} 0,$$

in probability.

We will follow the same arguments as for the proof of the uniform deterministic equivalent of the empirical spectral measure and the largest eigenvalue of Wigner matrices. We will begin by showing that the deterministic equivalent (5.46) we propose, holds uniformly in expectation. This is the object of the following lemma.

5.6.2 Lemma. *Let $\alpha \in (0,1)$. Let $X = (X_v)_{v \in \mathbb{Z}_+^d}$ be a family of i.i.d non-negative random variables with common distribution function satisfying (5.9). For any $r > 0$,*

$$\sup_{||H||_{\ell^\alpha} \le r} |\mathbb{E}T(X + nH) - \mathcal{T}_n(H)| \underset{n \to +\infty}{\longrightarrow} 0,$$

where $\mathcal{T}_n(H)$ is as in (5.46).

Proof. Let $\mathcal{A}_m$ denote the subset of vectors of $\mathcal{A}$ of size less or equal than m, and define $\hat{\mathcal{T}}_n^{(m)}$ by,

$$\hat{\mathcal{T}}_n^{(m)}(H) = \sup_{V \in \mathcal{A}_m} \Big\{ \sum_{i=0}^{p} H_{v_i}^+ + \sum_{i=0}^{p-1} \frac{1}{n} \mathbb{E}T_{v_i, v_{i+1}}(X) \Big\},$$

and $\mathcal{T}_n^{(m)}$,

$$\mathcal{T}_n^{(m)}(H) = \sup_{V \in \mathcal{A}_m} \Big\{ \sum_{i=0}^{p} H_{v_i}^+ + \sum_{i=0}^{p-1} g\Big(\frac{v_{i+1} - v_i}{n}\Big) \Big\},$$

where $V = (v_0, ..., v_p)$ for some $p \le m$. We begin by proving that there is some constant $C > 0$ depending on α, such that for any $||H||_{\ell^\alpha} \le r$,

$$-Cr(\log n)^{\frac{1}{\alpha}} n^{\alpha-1} \le \frac{1}{n}\mathbb{E}T(X+nH) - \hat{\mathcal{T}}_n^{(m)}(H) \le Crm^{1-\frac{1}{\alpha}}. \tag{5.47}$$

In the following C will denote a constant which will depend only on α and which will vary along the lines of the proof. Let π be an optimal path for the last-passage time $T(X+nH)$, and denote by $v_1, .., v_{m-1}$ be the $m-1$ largest values of H on the path π, sorted in lexicographic order. Add $v_0 = (0, ..., 0)$ and $v_m = (n, ..., n)$, to get $V = \{v_0, ..., v_m\} \in \mathcal{A}_m$. We have

$$\frac{1}{n}T(X+nH) - \sum_{i=0}^{m} H_{v_i}^+ - \sum_{i=0}^{m-1} \frac{1}{n} T_{v_i,v_{i+1}}(X) \le \frac{1}{n}\sum_{v\in\pi}(X+nH)_v^+ - \sum_{i=0}^{m} H_{v_i}^+ - \frac{1}{n}\sum_{v\in\pi} X_v.$$

As $(x+y)^+ \le x^+ + y^+$, we deduce

$$\frac{1}{n}T(X+nH) - \sum_{i=0}^{m} H_{v_i}^+ - \sum_{i=0}^{m-1} \frac{1}{n} T_{v_i,v_{i+1}}(X) \le \sum_{v\in\pi\cap V^c} H_v^+.$$

Now observe that if $M_1 \ge ... \ge M_{d(n+1)}$ are the values of H^+ (or H^-) along π in decreasing order, we have since $\sum_i M_i^\alpha \le r^\alpha$, for any $k \in \{1, ..., d(n+1)\}$,

$$M_k \le rk^{-1/\alpha}. \tag{5.48}$$

Therefore,

$$\sum_{v\in\pi\cap V^c} H_v^+ \le r \sum_{k=m-1}^{+\infty} k^{-1/\alpha} \le Crm^{1-\frac{1}{\alpha}},$$

for some constant $C > 0$. This proves the upper bound of (5.47). On the other hand, let $V = \{v_0, ..., v_p\} \in \mathcal{A}_m$. Considering the optimal paths from v_i to v_{i+1} in the last-passage time $T_{v_i,v_{i+1}}(X)$, for $i = 0, ..., p-1$ and their concatenation, we get,

$$\sum_{i=0}^{p} H_{v_i}^+ + \frac{1}{n}\sum_{i=0}^{p-1} T_{v_i,v_{i+1}}(X) - T(X+nH) \le \sum_{X_v \ge -nH_v} H_v^- + \sum_{X_v \le -nH_v} \frac{X_v}{n}.$$

Turning our attention to the first sum, we deduce by bounding the first n^α largest weights H_v^- by X_v/n, and using the bound (5.48) for the rest of the terms,

$$\mathbb{E}\Big(\sum_{X_v \ge -nH_v} H_v^-\Big) \le \frac{n^\alpha}{n}\mathbb{E}\sup_v X_v + r\sum_{k>n^\alpha} k^{-\frac{1}{\alpha}}.$$

From (5.21) and Jensen's inequality we have,

$$\mathbb{E}\sup_v X_v \le c(\log n)^{\frac{1}{\alpha}},$$

for some constant $c > 0$. We thus proved,

$$\mathbb{E}\Big(\sum_{X_v \ge -nH_v} H_v^-\Big) \le Cr(\log n)^{\frac{1}{\alpha}} n^{\alpha-1}.$$

On the other hand,

$$\mathbb{E}\Big(\sum_{X_v\le -nH_v}\frac{X_v}{n}\Big)\le\frac{1}{n}\mathbb{E}\big(X_0|\{v: X_0\le -nH_v\}|\big).$$

But $||H||_{\ell^\alpha}\le r$, thus

$$|\{v: X_0\le -nH_v\}|\Big(\frac{X_0}{n}\Big)^\alpha\le r.$$

Therefore,

$$\mathbb{E}\Big(\sum_{X_v\le -nH_v}\frac{X_v}{n}\Big)\le n^{\alpha-1}r\mathbb{E}X_0^{1-\alpha}.$$

which concludes the proof of the lower bound of (5.47). Comparing $\mathcal{T}_n^{(m)}$ and $\hat{\mathcal{T}}_n^{(m)}$, we get using the translation invariance in law (by vectors of $\mathbb{Z}_+^d$) of $(X_v)_{v\in\mathbb{Z}_+^d}$,

$$|\mathcal{T}_n^{(m)}(H)-\hat{\mathcal{T}}_n^{(m)}|\le m\max_{v\in\{0,\dots,n\}^d}\Big|\frac{1}{n}\mathbb{E}T_{0,v}(X)-g\Big(\frac{v}{n}\Big)\Big|.$$

As $\mathbb{E}T_{0,\lfloor nw\rfloor}(X)$ is coordinate-wise non-decreasing as a function of $w\in\mathbb{R}_+^2$, and converges to $g(w)$ which is continuous by [75, Theorem 2.3], we deduce that $w\mapsto\mathbb{E}T_{0,\lfloor nw\rfloor}(X)$ converges uniformly to g on $[0,1]^2$ by Dini's Theorem. Thus,

$$|\mathcal{T}_n^{(m)}(H)-\hat{\mathcal{T}}_n^{(m)}|\le m\varepsilon(n), \tag{5.49}$$

where $\varepsilon(n)\to+\infty$ when $n\to+\infty$.

Now, using the same argument as for the upper bound of (5.47), we see that

$$|\mathcal{T}_n^{(m)}(H)-\mathcal{T}_n(H)|\le Crm^{1-\frac{1}{\alpha}}, \tag{5.50}$$

for any $||H||_{\ell^\alpha}\le r$. Indeed, if V achieves the supremum in $\mathcal{T}_n(H)$, then taking V' the m largest values of H^+ on V, we get

$$0\le\mathcal{T}_n(H)-\mathcal{T}_n^{(m)}(H)\le\sum_{v\notin V'}H_v^+.$$

Thus, using (5.48), we get the claim. To summarize, we got by (5.47), (5.49), and (5.50),

$$\Big|\frac{1}{n}\mathbb{E}T(X+nH)-\mathcal{T}_n(H)\Big|\le Crm^{1-\frac{1}{\alpha}}+m\varepsilon(n)+Cr(\log n)^{\frac{1}{\alpha}}n^{\alpha-1},$$

for some constant $C>0$ and for any $||H||_{\ell^\alpha}\le r$, which gives finally the claim by taking the lim sup as $n\to+\infty$, and $m\to+\infty$.

□

Proof of Proposition 5.6.1. Let $H\in\mathbb{R}^I$. Note that $X\mapsto T(X+nH)$ is 1-Lipschitz with respect to $||\ ||_{\ell^1}$ on $\mathbb{R}^I$. As $||\ ||_{\ell^1}\le||\ ||_{\ell^\alpha}$ since $\alpha<1$, we deduce that $X\mapsto T(X+nH)$ is also 1-Lipschitz with respect to $||\ ||_{\ell^\alpha}$. Moreover by Hölder's inequality, $X\mapsto T(X+nH)$ is $\sqrt{n}$-Lipschitz with respect to $||\ ||_{\ell^2}$. We get by Lemma 2.2.6, for any $t>0$,

$$\mathbb{P}(|T(X+nH)-m|>tn)\le 8\exp(-cp_\alpha(t)),$$

where m is the median of $T(X+nH)$, c is some strictly positive constant, and

$$p_\alpha(t) = \min\Big(\frac{t^2 n}{(\log n)^{2(\frac{1}{\alpha}-1)}}, \frac{tn}{(\log n)^{\frac{1}{\alpha}-1}}, n^\alpha t^\alpha\Big).$$

Integrating this inequality we get,

$$|\mathbb{E}T(X+nH) - m| = O((\log n)^{\frac{1}{\alpha}-1} n^{\frac{1}{2}}),$$

uniformly in H. Using the result of Proposition 5.6.1, we deduce that for n large enough,

$$\mathbb{P}\big(|T(X+nH) - \mathcal{T}_n(H)| > (t+\delta_n)n\big) \leq 8e^{-cn^\alpha t^\alpha}, \tag{5.51}$$

where $\delta_n = O((\log n)^{\frac{1}{\alpha}-1} n^{-\frac{1}{2}})$. Let now $r > 0$. Note that

$$H \mapsto n^{-1}|T(X+nH) - \mathcal{T}_n(H)|,$$

is 2-Lipschitz with respect to $|| \,||_{\ell^1}$ on $\mathbb{R}^I$. Besides, by Lemma 5.5.10 for any $\varepsilon > 0$, the covering number of rB_{ℓ^α} by ℓ^2-balls of radii ε satisfies,

$$\log N(rB_{\ell^\alpha}, \varepsilon B_{\ell^2}) = O(\log n).$$

Since this estimate is negligible with respect to the concentration bound (5.51), we deduce using an ε-net arguments as in the proofs of Propositions 5.5.1 and 5.5.3, that

$$\mathbb{P}\Big(\sup_{H \in rB_{\ell^\alpha}} \big|\frac{1}{n}T(X+nH) - \mathcal{T}_n(H)\big| > t\Big) \underset{n\to+\infty}{\longrightarrow} 0,$$

which ends the proof of the claim. □

5.7 Applications to Wigner matrices

We apply in this section Theorem 5.2.1 to derive the LDP of Theorems 5.2.5, 5.2.7 and 5.2.9. In all this section, X will designate a Wigner matrix with the class $\mathcal{S}_\alpha$ for some $\alpha \in (0,2]$. It is clear that Theorem 5.2.1 remains valid in the context of Wigner matrices in the class $\mathcal{S}_\alpha$, making the according change in the rate function I_α, by replacing the weight function $|| \,||_{\ell^\alpha}^\alpha$ by W_α, which defines the law of a Wigner matrix in $\mathcal{S}_\alpha$ (see (5.8)).

5.7.1 Large deviations of the spectral measure

Proof of Theorem 5.2.5. From Proposition 5.40, we know that assumption (i) of Theorem 5.2.1 is satisfied with

$$\forall H \in \mathcal{H}_n,\ F_m(H) = \mu_{sc} \boxplus \mu_{n^{1/\alpha}H},$$

and

$$\forall H \in \mathcal{H}_n^{(\beta)},\ f_m(X) = \mu_{X/\sqrt{n}+n^{1/\alpha}H}.$$

where m is the (real) dimension of $\mathcal{H}_n^{(\beta)}$, with the distance d on $\mathcal{P}(\mathbb{R})$ defined in (2.12) and $v(m) = n^{1+\frac{\alpha}{2}}$.

By Lemma (2.4.1), we see that f_m is n^{-1} -Lipschitz with respect to $|| \, ||_{\ell^2}$ on $\mathcal{H}_n^{(\beta)}$ and d on $\mathcal{P}(\mathbb{R})$. By the remark 5.2.2 (c), and from the fact that $\alpha < 2$, we deduce that the assumption (ii) of Theorem 5.2.1 holds. Besides, as $\alpha \leq 2$, we have by [104, Theorem 3.32]

$$\forall H \in \mathcal{H}_n^{(\beta)}, \ (\mathrm{tr}|H|^{\alpha})^{1/\alpha} \leq ||H||_{\ell^\alpha}.$$

Thus for any $r > 0$,

$$F_m(rB_{\ell^\alpha}) \subset \{\mu \in \mathcal{P}(\mathbb{R}) : \mu|x|^\alpha \leq r^\alpha\},$$

which shows that $\cup_m F_m(rB_{\ell^\alpha})$ is relatively compact by Prokhorov's theorem, and that (iii) is verified.

To prove (iv) it is sufficient to show that for a fixed $H \in \mathcal{H}_p^{(\beta)}$, there is a sequence $H_n \in \mathcal{H}_n^{(\beta)}$, $n \geq p$, such that

$$\lim_{n\to+\infty} \mu_{n^{1/\alpha}H_n} = \mu_{p^{1/\alpha}H_p}, \quad \text{and} \quad \lim_{n\to+\infty} W_\alpha(H_n) = W_\alpha(H). \tag{5.52}$$

Let for any $k \in \mathbb{N}$, $H_{kp} = \oplus_{i=1}^k k^{-1/\alpha} H \in \mathcal{H}_{kp}^{(\beta)}$. We have $W_\alpha(H_{kp}) = W_\alpha(H)$, as $W_\alpha(\lambda Y) = \lambda^\alpha W_\alpha(Y)$ for any $\lambda > 0$, and

$$\mu_{(kp)^{1/\alpha}H_{kp}} = \mu_{p^{1/\alpha}H}.$$

Now, if $n = kp + l$, with $k \in \mathbb{N}$ and $1 \leq l \leq p$, we define

$$H_n = \Big(\frac{kp}{kp+l}\Big)^{1/\alpha} \begin{pmatrix} H_{kp} & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{H}_n^{(\beta)}.$$

We have,

$$\mu_{n^{1/\alpha}H_n} = \frac{kp}{kp+l}\mu_{(kp)^{1/\alpha}H_{kp}} + \frac{l}{kp+l}\delta_0.$$

Thus,

$$d(\mu_{n^{1/\alpha}H_n}, \mu_{(kp)^{1/\alpha}H_{kp}}) \leq \frac{2l}{kp+l} \leq \frac{2p}{n},$$

where d is defined in (2.12). Besides,

$$W_\alpha(H_{kp}) \geq W_\alpha(H_n) \geq \big(1 + \frac{1}{k}\big)^{-1/\alpha} W_\alpha(H_{kp}).$$

As $W_\alpha(H_{kp}) = W_\alpha(H)$, and $\mu_{(kp)^{1/\alpha}H_{kp}} = \mu_{p^{1/\alpha}H}$, we get the claim (5.52).

□

5.7.2 Large deviations of the largest eigenvalue

Proof of Theorem 5.2.7. We begin by giving back to J_α its variational form. We claim that for any $x \in \mathbb{R}$,

$$J_\alpha(x) = \sup_{\delta>0} \inf \big\{W_\alpha(A) : A \in \cup_{n\in\mathbb{N}} \mathcal{H}_n^{(\beta)}, \ |x - \rho(\lambda_A)| < \delta\big\}, \tag{5.53}$$

where ρ is the function

$$\forall x \in \mathbb{R},\ \rho(x) = \begin{cases} x + \frac{1}{x} & \text{if } x \geq 1, \\ 2 & \text{otherwise.} \end{cases}$$

Let us prove first that

$$\forall x \in \mathbb{R},\ J_\alpha(x) = \inf\left\{W_\alpha(A) : A \in \cup_{n\in\mathbb{N}}\mathcal{H}_n^{(\beta)},\ x = \rho(\lambda_A)\right\}. \tag{5.54}$$

When $x < 2$, both sides of (5.54) are infinite. If $x \geq 2$, we denote by $\mathcal{J}_\alpha$ the right-hand side of (5.54). Then, as $x \in (0,1] \mapsto \rho(1/x)$ is the inverse of the Stieltjes transform of μ_{sc} on $[2,+\infty)$ by (1.5), we can write

$$\mathcal{J}_\alpha(x) = \inf\left\{W_\alpha(A) : A \in \cup_{n\in\mathbb{N}}\mathcal{H}_n^{(\beta)},\ 1/\lambda_A = g_{\mu_{sc}}(x)\right\}.$$

As W_α is α-homogeneous, and $\lambda_{tA} = t\lambda_A$, for any $t \geq 0$, we get

$$\mathcal{J}_\alpha(x) = \mathcal{J}_\alpha(1) g_{\mu_{sc}}(x)^{-\alpha}.$$

Thus, $J_\alpha = \mathcal{J}_\alpha$. As J_α is clearly lower semi-continuous, the equality (5.53) holds by the remark 5.2.2(e) we made after Theorem 5.2.1.

We check now the assumptions of Theorem 5.2.1. Assumption (i) of Theorem 5.2.1 is met by the result of Proposition 5.5.3, with

$$\forall H \in \mathcal{H}_n,\ f_m(H) = \lambda_{X/\sqrt{n}},\ F_m(H) = \rho(\lambda_H),$$

where as before m is the dimension of $\mathcal{H}_n^{(\beta)}$, and $v(m) = n^{\alpha/2}$. Weyl's inequality (see (1.6)) shows that f_m is $n^{-1/2}$-Lipschitz with respect to $||\ ||_{\ell^2}$, and thus assumption (ii) is satisfied as $\alpha < 2$ by the remark 5.2.2 (c). Besides, note that as

$$|\lambda_H| \leq (\mathrm{tr}|H|^\alpha)^{1/\alpha} \leq ||H||_{\ell^\alpha},$$

for any $H \in \mathcal{H}_n^{(\beta)}$, and as ρ is non-decreasing, we deduce for any $r > 0$ that,

$$\{F_m(H) : H \in rB_{\ell^\alpha}\} \subset [2, \rho(r)],$$

which proves that (iii) is satisfied. To show that (iv) holds, it suffices to observe that if $H \in \mathcal{H}_n^{(\beta)}$, and if we set for any $m \geq n$,

$$H_m = \begin{pmatrix} H_n & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{H}_m^{(\beta)}, \tag{5.55}$$

then $W_\alpha(H_m) = W_\alpha(H)$, and provided $\lambda_H \geq 0$, we have $\lambda_H = \lambda_{H_m}$, so that in particular $\rho(\lambda_H) = \rho(\lambda_{H_m})$. □

5.7.3 Large deviations of non-commutative polynomials

Finally, we give a proof of Theorem 5.2.9.

Proof of Theorem 5.2.9. By a homogeneity argument similar as for the proof of Theorem 5.2.7, we get for any $x \in \mathbb{R}$,

$$K_\alpha(x) = \inf\left\{W_\alpha(\mathbf{H}) : \mathbf{H} \in \cup_{n\in\mathbb{N}}(\mathcal{H}_n^{(\beta)})^p, x = \mathrm{tr}P_d(\mathbf{H}) + \tau(P(\mathbf{s}))\right\},$$

where P_d denotes the homogeneous part of degree d of P. From the remark 5.2.2 (e), we get as K_α is lower semi-continuous, that

$$K_\alpha(x) = \sup_{\delta>0}\inf\left\{W_\alpha(\mathbf{H}) : \mathbf{H} \in \cup_{n\in\mathbb{N}}(\mathcal{H}_n^{(\beta)})^p, |x - \mathrm{tr}P_d(\mathbf{H}) - \tau(P(\mathbf{s}))| < \delta\right\}.$$

Assumption (i) of Theorem 5.2.1 is a consequence of Lemma 5.5.4 with the speed $v(m) = n^{\alpha(\frac{1}{2}+\frac{1}{d})}$ and

$$F_m(\mathbf{H}) = \mathrm{tr}P_d(\mathbf{H}) + \tau(P(\mathbf{s})),\ f_m(\mathbf{H}) = \tau_n(P(\mathbf{X}/\sqrt{n})),$$

where m is the real dimension of $(\mathcal{H}_n^{(\beta)})^p$.

Let us now prove assumption (ii). Note that by linearity, it suffices to prove assumption (ii) when P is a monomial of total degree $k \geq 1$ less or equal than d, which we will assume from now on. If $k = 1$, then there are two cases to consider. First we see by Hölder's inequality that f_m is n^{-1}-Lipschitz with respect to $||\ ||_{\ell^2}$. If $d = 1$ then $\alpha \in (0, 1)$, so that as $v(n) = n^{3\alpha/2}$ in this case. We conclude by remark 5.2.2 (c) that assumption (ii) holds. If $d \geq 2$ and $k = 1$, then we deduce again by remark 5.2.2 (c) that assumption (ii) is fulfilled as $v(n) = n^{\alpha(\frac{1}{2}+\frac{1}{d})}$.

Assume now $k \geq 2$. Let $\mathbf{K} \in rB_{\ell^\alpha}$, and set $\mathbf{Y} = \mathbf{X} + n^{\frac{1}{2}+\frac{1}{d}}\mathbf{K}$. As we assumed P is a monomial of total degree k, from Lemma 2.6.4, we have for any $\mathbf{H} \in (\mathcal{H}_n^{(\beta)})^p$,

$$|f_m(\mathbf{Y}+\mathbf{H}) - f_m(\mathbf{Y})| \leq \frac{c}{n}\left(||\mathbf{Y}/\sqrt{n}||_{2(k-1)}^{k-1} + ||\mathbf{H}/\sqrt{n}||_2^{k-1}\right)||\mathbf{H}/\sqrt{n}||_2.$$

where c is some constant depending p and d. Using the fact that $x^{k-1} \leq 1 + x^{d-1}$ for any $1 \leq k \leq d$ and $x \geq 0$, we get,

$$\begin{aligned}|f_m(\mathbf{Y}+\mathbf{H}) - f_m(\mathbf{Y})| \leq{}& \frac{c}{n}(||\mathbf{Y}/\sqrt{n}||_{2(k-1)}^{k-1} + 1)||\mathbf{H}/\sqrt{n}||_2 \\ &+ \frac{c}{n}||\mathbf{H}/\sqrt{n}||_2^d.\end{aligned}$$

Let $\delta \in (0, 1)$ and $t_\delta = \delta n^{\frac{1}{2}+\frac{1}{d}}$. For $\mathbf{H} \in t_\delta B_{\ell^2}$,

$$|f_m(\mathbf{Y}+\mathbf{H}) - f_m(\mathbf{Y})| \leq 2c\delta(n^{\frac{1}{d}-1}||\mathbf{Y}/\sqrt{n}||_{2(k-1)}^{k-1} + 1).$$

With the notation of Theorem 5.2.1, we have

$$\mathbb{E}\sup_{\mathbf{H}\in t_\delta B_{\ell^2}} \mathcal{L}_m(\mathbf{H}) \leq 2c\delta(n^{\frac{1}{d}-1}\mathbb{E}||\mathbf{Y}/\sqrt{n}||_{2(k-1)}^{k-1} + 1),$$

where m is the dimension of $(\mathcal{H}_n^{(\beta)})^p$. By convexity, we deduce

$$\mathbb{E}||\mathbf{Y}/\sqrt{n}||_{2(k-1)}^{k-1} \leq 2^{k-2}\mathbb{E}||\mathbf{X}/\sqrt{n}||_{2(k-1)}^{k-1} + 2^{k-2}n^{\frac{k-1}{d}}||\mathbf{K}||_{2(k-1)}^{k-1}.$$

But by Wigner's theorem,

$$\mathbb{E}||\mathbf{X}/\sqrt{n}||_{2(k-1)}^{k-1} \leq c_0 n^{1/2},$$

for some constant $c_0 > 0$. As $\mathbf{K} \in rB_{\ell^\alpha}$ with $\alpha \leq 2$, we deduce as $k \geq 2$,

$$||\mathbf{K}||_{2(k-1)} \leq ||\mathbf{K}||_2 \leq r.$$

Thus,

$$\mathbb{E} \sup_{\mathbf{H} \in t_\delta B_{\ell^2}} \mathcal{L}_m(\mathbf{H}) \leq C\delta(n^{\frac{1}{d}+\frac{1}{2}-1} + r^{d-1}).$$

where C is some positive constant depending on p and d. This shows that assumption (ii) is satisfied.

We show now that assumption (iii) holds. Using (5.45) for $q = d$, we get,

$$|\mathrm{tr} P_d(\mathbf{H})| \leq C||\mathbf{H}||_\alpha^{d/\alpha},$$

for some constant $C > 0$. This proves condition (iii) of Theorem 5.2.1. To show that the last assumption (iv) is met, it suffices to observe that for any fixed $\mathbf{H} \in (\mathcal{H}_n^{(\beta)})^p$, with the same construction as in (5.55), there is a sequence $\mathbf{H}_m \in (\mathcal{H}_m^{(\beta)})^p$, for $m \geq n$, such that

$$\mathrm{tr} P_d(\mathbf{H}_m) = \mathrm{tr} P_d(\mathbf{H}_n),$$

and $W_\alpha(\mathbf{H}_n) = W_\alpha(\mathbf{H}_m)$.

$\square$

5.8 Application to last-passage time

We prove in this last section Theorem 5.2.11.

Proof of Theorem 5.2.11. We will verify the assumptions of Theorem 5.2.3. Assumption (i) holds due to Proposition 5.6.1 with $v(n) = n^\alpha$, and

$$\forall X \in \mathbb{R}^I, \; f_m(X) = \frac{1}{n} T(X^+), \; F_m(X) = \mathcal{T}_n(X),$$

where $\mathcal{T}_n$ is defined in (5.46), X^+ denotes the matrix with coefficients $(X_v^+)_v$, and m is the dimension of $\mathbb{R}^I$. As

$$X \mapsto T(X^+)/n,$$

is $n^{-1/2}$-Lipschitz with respect to $|| \; ||_{\ell^2}$, assumption (ii) is satisfied by the remark 5.2.2 (c).

Using the fact that $|| \; ||_{\ell^1} \leq || \; ||_{\ell^\alpha}$ when $\alpha \leq 1$, on $\mathbb{R}^I$, we see that the condition (iii) of Theorem 5.2.3 is met. To prove $(iv)'$, we first observe that

$$L_\alpha(x) = \inf\{||H||_{\ell^\alpha}^\alpha : \mathcal{T}_n(H) = x, H \in \mathbb{R}^I\}, \tag{5.56}$$

where $\mathcal{T}_n$ is defined in (5.46). Indeed, since the function g is superadditive by [75, Proposition 2.1], we deduce that

$$\mathcal{T}_n(H) \geq g(1, ..., 1),$$

for any $H \in \mathbb{R}^I$. Therefore, both sides of (5.56) are infinite if $x < g(1, ..., 1)$. Now if $x \geq g(1, ..., 1)$, and $H \in \mathbb{R}^I$ is such that $\mathcal{T}_n(H) = x$, then denoting $\{v_0, ..., v_p\}$ the element of $\mathcal{A}_m$ achieving the supremum in (5.46), we get,

$$||H||_{\ell^\alpha}^\alpha \geq \Big(\sum_{i=0}^{p-1} H_{v_i}^+ \Big)^\alpha = \Big(x - \sum_{i=0}^{p-1} g\Big(\frac{v_{i+1} - v_i}{n} \Big) \Big)^\alpha.$$

Using the superadditivity of g, it yields

$$||H||_{\ell^\alpha}^\alpha \geq (x - g(1, ..., 1))^\alpha,$$

with equality for the matrix H whose entries are all zero except $H_{n,n} = x - g(1, ..., 1)$. This proves the equality (5.56). In particular, L_α is lower semi-continuous and therefore by the remark 5.2.2 (e), we deduce,

$$L_\alpha(x) = \sup_{\delta > 0} \inf\{||H||_{\ell^\alpha}^\alpha : |\mathcal{T}_n(H) - x| < \delta, H \in \mathbb{R}^I)\}.$$

As the matrices $H \in \mathcal{M}_{n+1}(\mathbb{R})$ with $H_v = (x - g(1, ..., 1))_+ \mathbb{1}_{v=(n,n)}$, achieves (5.56) for any n, we deduce,

$$L_\alpha(x) = \sup_{\delta > 0} \limsup_{n \to +\infty} \inf\{||H||_{\ell^\alpha}^\alpha : |\mathcal{T}_n(H) - x| < \delta, H \in \mathbb{R}^I)\}.$$

Finally, as $\mathcal{T}_n(H) = \mathcal{T}_n(H^+)$, where H^+ is the matrix $(H_v^+)_{v \in \{0,...,n\}^2}$, we get

$$L_\alpha(x) = \sup_{\delta > 0} \limsup_{n \to +\infty} \inf\{||H||_{\ell^\alpha}^\alpha : |\mathcal{T}_n(H) - x| < \delta, H \in \mathbb{R}_+^I\}.$$

This proves the last assumption $(iv)'$ of Theorem 5.2.3.

□

5.9 Conclusion and perspectives

We showed that the approach of Borell and Ledoux for the large deviations of Wiener chaoses can, in some extent, give another interpretation of the heavy-tail phenomena appearing in large deviations. The limitation of Theorem 5.2.1 to the probability measures ν_α^n is mainly technical. It stems from the fact that they are the only probability measures for which we know how to make the upper bounds (5.7) and lower bound (5.6) match for the same weight function c_α.

Concerning the upper bound, as we mentioned in the introduction, we do not know how to get it if we break the symmetry and consider the measure μ_α instead of ν_α for example. More generally, it seems -to our knowledge- quite an open question to get an optimal weight function which satisfies, for a given probability measure on $\mathbb{R}^n$, the τ-property. Concerning product measures, we mention in this direction the work of Latała and Wojtaszczyk [66, Theorem 2.18] who showed that any product log-concave probability measure on $\mathbb{R}^n$ satisfies the τ-property with weight function the Legendre transform of the Log-Laplace transform (up to a scaling factor). Still, one can show that if $\mu = Z^{-1} e^{-V} dx$ is a log-concave probability measure on $\mathbb{R}$ such that $V(x) \sim_{\pm\infty} |x|^\alpha$ for some $\alpha \in (0, 2]$, then the same estimates as for the Brenier map from ν_1 to ν_α hold for the Brenier map from ν_1 to μ. Therefore, the upper bound (5.7) must be true with μ^n instead of ν_α^n and with $c_\alpha = || \ ||_\alpha^\alpha$.

The weak point of Theorem 5.2.1 is the lower bound (5.6), because it somewhat implies that the potential of the probability measure we consider is homogeneous. Indeed, if μ is as above, with V symmetric, convex, differentiable and V' concave if $\alpha \in [1, 2]$, or concave and non-decreasing on $[0, +\infty)$ is $\alpha \in (0, 1)$, then we have, assuming $V(0) = 0$ (which we can always do),

$$V(x + y) \leq V(x) + V(y),$$

if $\alpha \in (0, 1)$, and

$$V(x + y) \leq V(x) + \mathrm{sg}(xy)V'(|x|)|y| + V(y),$$

if $\alpha \in [1, 2]$. The same proof as in Proposition 5.4.4 for $V = |x|^\alpha$ would yields

$$\mu^n(E + v(n)^{1/\alpha}h) \geq e^{-\sum_{i=1}^n V(v(n)^{1/\alpha}h_i)+o(v(n))}.$$

Unfortunately, this lower bound at the exponential scale $v(n)$ can be very different from $||h||_{\ell^\alpha}^\alpha$. Thus, the lower bound of Proposition 5.4.4 is esssentially only valid for ν_α.

More daringly, one can hope a possible extension of Theorem 5.2.1 in the case of the symmetric exponential measure to large deviations which do not rely on heavy-tail phenomena. Indeed, Talagrand's transportation-cost inequality for the symmetric exponential measure [94, Theorem 1.2] has the striking feature that the weight function c_λ (see Proposition 5.3.3) he provides, is optimal for *any* λ, meaning that for each λ there is a measure μ which realizes the equality in the transport inequality with cost c_λ. Thus, as this family of weights c_λ is truely optimal, one can wish for Theorem 5.2.1 to be true without assumption (ii) (similarly as in the Gaussian case), and with a more intricate rate function.

6. Large deviations of β-ensembles with quadratic potential

6.1 Introduction

In this last chapter, we make an incursion outside heavy-tail phenomena in the large deviations of random matrices. Our primary motivation in this chapter is to give another proof of the large deviations of the GUE and GOE than the one of Ben Arous and Guionnet [4], which does not rely on the knowledge of the law of the spectrum.

To this end, we come back to the β-ensemble with quadratic potential, defined as the probability measure $\mathbb{P}_\beta^n$ on $\mathbb{R}^n$, by

$$d\mathbb{P}_\beta^n = \frac{1}{Z_\beta^n} \prod_{i<j} |\lambda_i - \lambda_j|^\beta e^{-\frac{n\beta}{4}\sum_{i=1}^n \lambda_i^2} \prod_{i=1}^n d\lambda_i.$$

We saw in the Introduction §1.7.2 that for $\beta = 1, 2$, $\mathbb{P}_\beta^n$ is the joint law of the eigenvalues of respectively, a (normalized) GOE and GUE matrix respectively, whereas for $\beta > 0$, we have the following tri-diagonal representation due to Dimitriu and Edelman [46].

Let $(\xi_i, \zeta_i)_{i\geq 1}$ be a collection of independent random variables, such that $(\xi_i)_{i\geq 1}$ are Gaussian variables with variance 2, and ζ_i follows a χ-distribution of degree $i\beta$. We recall that the χ-distribution of degree $d > 0$ is the probability measure on $\mathbb{R}_+$ with density f_d,

$$\forall x \in \mathbb{R}_+, \; f_d(x) = \frac{2^{1-d/2}}{\Gamma(d/2)} x^{d-1} e^{-x^2/2}.$$

We form J_n, the tridiagonal symmetric matrix of size n,

$$J_n = \frac{1}{\sqrt{\beta n}} \begin{pmatrix} \xi_1 & \zeta_1 & 0 & \cdots & 0 \\ \zeta_1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \zeta_{n-1} \\ 0 & \cdots & 0 & \zeta_{n-1} & \xi_n \end{pmatrix}. \tag{6.1}$$

The tridiagonalization procedure shows that a GUE or a GOE matrix is unitary or orthogonally conjugated to a Jacobi matrix with the same law as J_n. Classically, the tridiagonalization is done by beginning from the first coordinate vector so that one ends up with a Jacobi matrix equal in law to the "reversed" version of J_n, that is the matrix $M_\sigma J_n M_\sigma^{-1}$, where M_σ is the permutation matrix associated to $\sigma = (12...n)^{n-1}$. Performing the tridiagonalization from the last coordinate vector or using the invariance in law by the unitary or orthogonal group, one recovers our Jacobi matrix J_n. The reason for this choice will become clearer later.

We recall the following LDP result proved by Ben-Arous and Guionnet [4].

6.1.1 Theorem ([4, Theorem 1.3])**.** *The empirical spectral measure of J_n, denoted by μ_n, follows a LDP with speed n^2 and good rate function I_β, defined for any $\mu \in \mathcal{P}(\mathbb{R})$,*

$$I_\beta(\mu) = \frac{\beta}{2}\Big(\frac{1}{2}\int x^2 d\mu(x) - \Sigma(\mu) + \frac{1}{2}\log\beta - \frac{1}{4}\Big),$$

with $\Sigma(\mu)$ the non-commutative entropy of μ, that is,

$$\Sigma(\mu) = \begin{cases} \int \log|x-y| d\mu(x)d\mu(y) & \text{if } \int \log(1+|x|)d\mu(x) < +\infty, \\ -\infty & \text{otherwise.} \end{cases}$$

In this chapter we will provide another proof of this result using solely the tridiagonal representation of J_n. In particular, for $\beta = 1$ and 2, our proof of the LDP of the spectral measure of the associated classical Gaussian ensembles will only take advantage of the matricial form of those models.

6.2 Large deviations by the objective method

The general strategy is to prove the LDP by a contraction argument (see [43, Theorem 4.2.1]). Let $(X_k, Y_k)_{1\le k\le n}$ be the diagonal and off-diagonal coefficients of J_n, that is, for any $k \in \{1, ..., n\}$, $X_k = J_n(k,k)$ and $Y_k = J_n(k, k+1)$ with the convention that $Y_n = 0$. As we mentioned in the Introduction §1.7.3, one can associate to J_n the law of an infinite Jacobi matrix with stationary coefficients, by considering the law of the Jacobi matrix with diagonal and off-diagonal coefficients $(X_{T+k[n]}, Y_{T+k[n]})_{k\in\mathbb{Z}}$, where T is uniformly sampled in $\{1, ..., n\}$, and $[n]$ stands for "mod n". This defines a stationary probability measure ρ_n on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$.

Let $\mathcal{P}_{\text{stat}}$ be the set of stationary probability measures on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$, equipped with the Borel σ-algebra of the product topology. One can associate to an element $\rho \in \mathcal{P}_{\text{stat}}$, seen as the law of an infinite Jacobi matrix with stationary coefficients, a natural *spectral measure*, μ_ρ defined by,

$$\mu_\rho = \mathbb{E}_\rho \mu_J^{e_o}, \tag{6.2}$$

where under $\mathbb{P}_\rho$, J has law ρ, and where $\mu_J^{e_o}$ denotes the spectral measure associated to the first coordinate vector e_o. The spectral measure of J at e_o is well defined because J is almost surely essentially self-adjoint. Indeed, let $(v_k, w_k)_{k\in\mathbb{Z}}$ denote the diagonal and off-diagonal coefficients of J. We have thus $w_k > 0$ for any $k \in \mathbb{Z}$.

J defines a discrete Schrödinger operator on $\ell^2_f(\mathbb{Z})$, the subspace of vectors of $\ell^2(\mathbb{Z})$ having a finite number of non-zero coordinates, by

$$\forall \psi \in \ell^2_f(\mathbb{Z}), \forall n \in \mathbb{Z},\ J\psi_n = v_n\psi_n + w_{n-1}\psi_{n-1} + w_n\psi_{n+1}.$$

We know by [36, Chapter III §III.2] that a sufficient condition for J to be essentially self-adjoint is that,

$$\sum_{k=0}^{+\infty} \frac{1}{w_k^2} = \sum_{k=1}^{+\infty} \frac{1}{w_{-k}^2} = +\infty. \tag{6.3}$$

If $\int w_o^2 d\rho < +\infty$, then by Jensen's inequality we see, as ρ is stationary,

$$\int \Big(\sum_{k=0}^{+\infty} \frac{1}{w_k^2} \Big) d\rho \geq \sum_{k=0}^{+\infty} \Big(\int w_k^2 d\rho \Big)^{-1} = +\infty,$$

and similarly for the coefficients on $\mathbb{Z}_-$. We deduce that the spectral measure of any $\rho \in \mathcal{P}_{\text{stat}}$ such that

$$\int w_o^2 d\rho < +\infty, \tag{6.4}$$

is well defined. Actually, J is almost surely under ρ, essentially self-adjoint due to more abstract arguments from the theory of affiliated operators to Von Neumann algebra, which is is far beyond the scope of this text. We refer the reader to [28] for an exposition of the argument. But as we will see, it will be sufficient for us here to define the spectral measure for $\rho \in \mathcal{P}_{\text{stat}}$ such that $\int w_o^2 d\rho < +\infty$.

If J is a deterministic Jacobi matrix of size n with coefficients $(b_k, a_k)_{1 \leq k \leq n}$, where by convention we set $a_n = 0$, that is,

$$J = \begin{pmatrix} b_1 & a_1 & 0 & \cdots & 0 \\ a_1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & a_{n-1} \\ 0 & \cdots & 0 & a_{n-1} & b_n \end{pmatrix},$$

then one can associate the law ρ of the infinite Jacobi matrix with stationary coefficients $(b_{k+T[n]}, a_{k+T[n]})_{k \in \mathbb{Z}}$, where T is uniformly sampled in $\{1, ..., n\}$. Then one can check that the spectral measure of ρ in the sense of (6.2) coincides with the empirical spectral measure of J.

A key argument in our approach is that the map which associates to a probability measure $\rho \in \mathcal{P}_{\text{stat}}$ its spectral measure, is continuous for the weak topology (see [28, Proposition 2.2]). For sake of completeness, we reproduce the argument of [28, Proposition 2.2] in our simpler setting.

6.2.1 Lemma. *Let*

$$\mathcal{E}_2 = \big\{ \rho \in \mathcal{P}_{stat} : \int w_o^2 d\rho < +\infty \big\}.$$

The map $\rho \in \mathcal{E}_2 \mapsto \mu_\rho$, where μ_ρ is defined in (6.2), is continuous for the weak topology.

Proof. Let ρ_n, $\rho \in \mathcal{E}_2$ such that ρ_n converges weakly to ρ. As $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$ is second countable, the support of ρ is second-countable and in particular separable. By Skorokhod's theorem (see [24, Theorem 6.6]), there are some infinite Jacobi matrices J_n, with law ρ_n and J with law ρ, defined on the same probability space, such that, almost surely, $J_n\psi$ converges to $J\psi$, for any finitely supported $\psi \in \ell^2_f(\mathbb{Z})$. By [84, Theorem VIII. 25 (a)]), this implies the convergence of the resolvents on $\ell^2_f(\mathbb{Z})$, and therefore the weak convergence of $\mu^{eo}_{J_n}$ to μ^{eo}_J almost surely. By dominated convergence, we deduce the weak convergence of μ_{ρ_n} to μ_ρ. $\square$

The continuity of the spectral measure allows us to carry out a contraction argument to derive Theorem 6.1.1, and to focus on deriving a LDP for J_n identified to ρ_n, the law of $(X_{T+k[n]}, Y_{T+k[n]})_{k\in\mathbb{Z}}$, where T is uniformly sampled in $\{1, ..., n\}$. Actually, to derive the LDP of ρ_n, we will need to enrich it of the data T/n, allowing us to keep track of the "observing point" T. Therefore, we define π_n to be the law of

$$((X_{T+k[n]}, Y_{T+k[n]})_{k\in\mathbb{Z}}, T/n), \tag{6.5}$$

where $(X_k, Y_k)_{k\in\mathbb{Z}}$ are the coefficients of J_n, defined in (6.1), and T is again uniformly sampled in $\{1, ..., n\}$.

6.3 A variational formula for the non-commutative entropy

We will actually derive two LDP. On one hand, we will derive the LDP for the empirical spectral measure of J_n, which will be given for the lower bound, by a localization argument similar to the one used in [4], and for the upper bound by the contraction of the upper bound we will derive for π_n, with a rate function formulated in terms of the spectral measure. Besides, we will also give a LDP for π_n with a rate function involving the coefficients. The contraction principle gives us two rate functions for μ_n. By uniqueness of the rate function (see [43, Lemma 4.1.4]), we get an equality which will give us a variational formula for the non-commutative entropy, which is stated in the next proposition. This approach is typically in the spirit of the proofs of the "sum rules" obtained by Gamboa, Nagel and Rouault, in [52] and [51] for example.

6.3.1 Proposition. *For any $\mu \in \mathcal{P}(\mathbb{R})$ such that $\mu(x^2) < +\infty$, we have the following equality in $\mathbb{R} \cup \{-\infty\}$,*

$$\int \log|x-y| d\mu(x)d\mu(y) = 2\sup_\pi \mathbb{E}_\pi T \log w_0, \tag{6.6}$$

where the supremum is taken on all couplings π between $\rho \in \mathcal{P}_{stat}$ such that $\mu_\rho = \mu$, and the uniform measure on $[0,1]$, and where under $\mathbb{P}_\pi$, $((v_k, w_k)_{k\in\mathbb{Z}}, T)$ has law π.

Before going into the proofs of Theorem 6.1.1 and Proposition 6.3.1, we make some comments on this last result. Note that if $\rho \in \mathcal{P}_{\mathrm{stat}}$ is such that $\mu_\rho = \mu$, with $\mu(x^2) < +\infty$, we get from the above proposition by taking the independent coupling,

$$\int \log|x-y|d\mu(x)d\mu(y) \geq \mathbb{E}_\rho \log w_0, \tag{6.7}$$

where under $\mathbb{P}_\rho$, $(v_k, w_k)_{k\in\mathbb{Z}}$ has law ρ. This inequality for ergodic Jacobi matrices is known as a consequence of the Thouless formula, which says in this setting (see [88, Theorem 7.1 (7.17)]), that almost surely,

$$\gamma(z) = \log(A^{-1}) + \int \log|z-x|d\mu(x),$$

where γ is the Lyapunov exponent, defined as the exponential growth of the norm of the transition matrices (see [88, section 7] for proper definitions), and

$$A = \lim_{n\to+\infty}(w_0 w_1 ... w_n)^{1/n}.$$

As ρ is ergodic, $\log A$ is almost surely deterministic and equal to $\mathbb{E}_\rho \log w_0$. Integrating the Thouless formula against μ, and using the fact that $\gamma \geq 0$ by [88, Theorem 7.1 (7.18)], yields (6.7).

One can wonder what kind of optimal coupling can arise from the formula of Proposition 6.3.1. We will identify on a simple example the optimal coupling. First note that if μ_w is the arcsine law with variance w, that is, the probability measure on $\mathbb{R}$,

$$\mu_w = \frac{1}{\pi\sqrt{4w^2 - x^2}}\mathbb{1}_{|x|<2w}dx,$$

then

$$\int \log|x-y|d\mu_w(x) = \log w, \tag{6.8}$$

for any $y \in [-2w, 2w]$, by a direct computation following the lines of [4, Lemma 2.27]. Note that μ_1 is the spectral measure at e_o, of the two-sided free Jacobi matrix J_{free} with coefficients $J_{\text{free}}(k,k) = 0$ and $J_{\text{free}}(k, k+1) = 1$ for any $k \in \mathbb{Z}$ (see [28, §1.5.2]). Now, if W is a positive random variable and J is the infinite Jacobi matrix with coefficients $v_k = 0$ and $w_k = W$, then the spectral measure μ, of the law of J, in the sense of (6.2), is the law of WX, where X is independent of W and has law μ_1. If (Y, W') is an independent copy of (X, W), then we can write

$$\mathbb{E}\log|WX - W'Y| = 2\mathbb{E}\mathbb{1}_{W\leq W'}\log|WX - W'Y|.$$

Then, conditioning on all the variables but X, we deduce by (6.8),

$$\int \log|x-y|d\mu(x)d\mu(y) = 2\mathbb{E}\mathbb{1}_{W'\leq W}\log W = 2\mathbb{E}F(W)\log W, \tag{6.9}$$

where F denotes the distribution function of W. Actually, on can also show that the same equality holds if we put on the diagonal of J a random variable V independent from W. This shows that the optimal coupling in (6.6) for μ, is given by the choice $T = F(W)$. One can apply this formula (6.9) to the semi-circular law μ_{sc}, as one can verify that $\mu_{sc} = \mathbb{E}\mu_{\sqrt{T}}$, where T is a random variable with uniform law on $[0,1]$. We get,

$$\int \log|x-y|d\mu_{sc}(x)d\mu_{sc}(y) = \mathbb{E}T\log T.$$

6.4 Large deviations of the β-ensembles

We will give in this section a proof of Theorem 6.1.1 based on a contraction argument of large deviations bounds of J_n.

Proof of the lower bound. As argued in the proof of Ben Arous and Guionnet [4], we can focus our efforts on deriving the lower bound for compactly supported measures such that $I_\beta(\mu) < +\infty$. Let μ be such a measure. In particular, this means $\Sigma(\mu) > -\infty$, therefore μ does not have atoms. Let F be the distribution function of μ and $F^{(-1)}$ its generalized inverse. We define for any $k \in \{1, ..., n\}$,

$$z_k^{(n)} = F^{(-1)}\big(\frac{k}{n+1}\big).$$

Let $P_n = \prod_{k=1}^n (x - z_k^{(n)})$. As μ has no atoms, the zeros of P_n as P_{n-1} are strictly interlacing. The Gerominus-Wendroff theorem [87, Chapter 1, 1.2 §6] tells us that there is a Jacobi matrix of size n, K_n such that

$$P_n = \det(x - K_n), \quad P_{n-1} = \det(x - K_n^{(n)}),$$

where $K_n^{(n-1)}$ denotes the $(n-1)$ principal sub-matrix of K_n. By definition of the $z_k^{(n)}$'s, the empirical spectral measure of K_n, denoted ν_n, converges weakly to μ. Let $\delta > 0$. We can write, for n large enough,

$$\mathbb{P}(\mu_n \in B(\mu, 2\delta)) \geq \mathbb{P}(\mu_n \in B(\nu_n, \delta))$$

By Hoffman-Wielandt inequality (1.2), we deduce,

$$\mathbb{P}\big(\mu_n \in B(\nu_n, \delta)\big) \geq \mathbb{P}\big(||J_n - K_n||_\infty < \delta/2\big). \tag{6.10}$$

Let $b_1, ..., b_n$ and $a_1, ..., a_{n-1}$ denote respectively the diagonal and off-diagonal entries of K_n. In order to get a lower bound on $\mathbb{P}\big(||J_n - K_n||_\infty < \delta/2\big)$, we will need the following standard estimates on the Gaussian measure and χ-distribution. As we want our estimates to hold uniformly, we take some care to write them precisely.

6.4.1 Lemma. *Let $\delta < \beta$. For any $b \in \mathbb{R}$, $a > 0$,*

$$\mathbb{P}(|\xi_1/\sqrt{\beta n} - b| < \delta) \leq e^{-\frac{n\beta}{4}(1-2\delta)b^2 + n\varepsilon_n(\delta)},$$

and for $i \in \{1, ..., n\}$,

$$\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) \leq \frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)}(a+\delta)^{i\beta} e^{-n(1-4\delta)(a+\delta)^2 + n\varepsilon_n(\delta)},$$

where $\varepsilon_n(\delta)$ is some function such that $\limsup_n \varepsilon_n(\delta) = o(1)$, uniformly in $a > 0$ and b. Let $R > 0$. For any $0 < a < R$, and $|b| < R$

$$\mathbb{P}(|\xi_i/\sqrt{\beta n} - b| < \delta) \geq e^{-\frac{n\beta}{4}b^2 - n\eta_n(\delta)},$$

and for $i \in \{1, ..., n\}$,

$$\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) \geq \frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} a^{i\beta} e^{-na^2 - n\eta_n(\delta)},$$

where $\limsup_n \eta_n(\delta) = o(1)$ is uniform in $0 < a < R$ and $|b| < R$.

Proof. Let $\delta < \beta$. For $i \in \{1, ..., n\}$, and $a > 0$,

$$\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) = \frac{2^{1-i\beta/2}(\beta n)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} \int_{\substack{a-\delta \leq x < a+\delta \\ x>0}} x^{i\beta-1} e^{-\frac{\beta n}{2}x^2} dx$$

We have,

$$\begin{aligned}\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) &\geq \frac{2(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} \int_{a \leq x < a+\delta} x^{i\beta-\delta} e^{-\frac{\beta n}{2}x^2} x^{-1+\delta} dx \\ &\geq \frac{2(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} a^{i\beta-\delta} e^{-\frac{\beta n}{2}(a+\delta)^2} \int_{0<x<\delta} x^{-1+\delta} dx.\end{aligned}$$

Therefore,

$$\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) \geq \frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} a^{i\beta} e^{-\frac{\beta n}{2}a^2 - n\eta_n(\delta)},$$

with $\limsup_n \eta_n(\delta) = o(1)$, uniformly in $|a| \leq R$. Similarly,

$$\begin{aligned}\mathbb{P}(|\zeta_i/\sqrt{\beta n} - a| < \delta) &\leq \frac{2(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} \int_{a-\delta \leq x < a+\delta} x^{i\beta-1} e^{-\frac{\beta n}{2}x^2} dx \\ &\leq \frac{4\delta(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)} (a+\delta)^{i\beta-1} e^{-\frac{\beta n}{2}(a-\delta)^2}\end{aligned}$$

As $(a+\delta)^{i\beta-1} \leq (a+\delta)^{i\beta}\delta^{-1}$, and $(a-\delta)^2 \geq (1-4\delta)(a+\delta)^2 + O(\delta)$, we get the claim. □

We now come back to the proof of the lower bound of Theorem 6.1.1. By (6.10) and Lemma 6.4.1, we get

$$\mathbb{P}\big(\mu_n \in B(\nu_n, \delta)\big) \geq e^{-n^2(I_n - c_{\beta,n}) - n^2\varepsilon_n(\delta)}, \tag{6.11}$$

where

$$I_n = \frac{\beta}{4n}\sum_{i=1}^{n}(b_i^2 + 2a_i^2) - \beta\sum_{i=1}^{n-1}\frac{i}{n}\log a_i, \tag{6.12}$$

and

$$c_{\beta,n} = \frac{1}{n^2}\log\prod_{i=1}^{n-1}\frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)},$$

and with $\limsup_n \varepsilon_n(\delta) = o(1)$. As the target measure μ is compactly supported, the $z_k^{(n)}$ are in a compact set. Thus the spectral radius of K_n is uniformly bounded, in particular as K_n is Hermitian, the coefficients of K_n are bounded. Therefore the error term ε_n in (6.11) can be made uniform in the coefficients of K_n.

We begin by estimating the contributions of the normalization constants at the exponential scale n^2,

$$\lim_{n\to+\infty}\frac{1}{n^2}\log\prod_{i=1}^{n-1}\frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)}.$$

Note that by Cesarò theorem, we can replace $\Gamma(i\beta/2)$ by its equivalent at $+\infty$ given by Stirling formula. This gives,

$$\lim_{n\to+\infty}\frac{1}{n^2}\log\prod_{i=1}^{n-1}\frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)}=\lim_{n\to\infty}\Big[\frac{\beta}{4}-\frac{1}{n}\sum_{i=1}^{n}\frac{i\beta}{2n}\log\big(\frac{i\beta}{n}\big)\Big].$$

But,

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{i}{n}\log\big(\frac{i}{n}\big)=\int_0^1 x\log x\ \mathrm{d}x=-\frac{1}{4}.$$

Thus,

$$\begin{aligned}\lim_{n\to+\infty}\frac{1}{n^2}\log\prod_{i=1}^{n-1}\frac{(n\beta)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)}&=\frac{\beta}{4}-\frac{\beta}{4}\log\beta-\frac{\beta}{8}\\&=\frac{\beta}{8}-\frac{\beta}{4}\log\beta=:c_\beta.\end{aligned}\tag{6.13}$$

Turning our attention to I_n, we recognize

$$\nu_n(x^2)=\frac{1}{n}\sum_{i=1}^{n}b_i^2+\frac{2}{n}\sum_{i=1}^{n-1}a_i^2.\tag{6.14}$$

To link the other term appearing in (6.12) with the non-commutative entropy, we prove the following lemma.

6.4.2 Lemma. *Let $J=(a_n,b_n)_{n\geq 1}$ be a Jacobi matrix. Let ν_n denote the empirical spectral measure of $J^{(n)}$, the principal sub-matrix of size n.*

$$\int\log|x-y|d\nu_n(x)d\nu_{n-1}(y)=\frac{2}{n(n-1)}\sum_{i=1}^{n-1}i\log a_i.$$

Proof. Let $(Q_n)_{n\in\mathbb{N}}$ be the monic orthogonal polynomials associated to J, that is, the polynomials which satisfy the three-terms recurrence relation

$$Q_{n+1}=(x-b_{n+1})Q_n-a_n^2Q_{n-1},$$

with the initial conditions $Q_{-1}=0$, $Q_0=1$ and where we set $a_0=1$. Equivalently, Q_n is the characteristic polynomial of $J^{(n)}$. Let $\mathrm{Res}(P,Q)$ denote the resultant of two polynomials P,Q. Recall that the resultant $\mathrm{Res}(P,Q)$ is defined as the determinant of a Sylvester matrix involving the coefficients of P and Q, and in particular

$$\mathrm{Res}(P,Q)=c_P^n\prod_{i,j}(\lambda_i-\lambda_j')=\prod_{j=1}^{n}P(\lambda_j')=\prod_{i=1}^{m}Q(\lambda_i),$$

where m and n are the respective degrees of P and Q, λ_i, λ_j' are the respective zeros of P and Q, and c_P is the coefficient of highest degree of P. From the three-terms recurrence relation, we get,

$$\mathrm{Res}(Q_{n+1},Q_n)=\mathrm{Res}(-a_n^2Q_{n-1},Q_n)=(-a_n^2)^n\mathrm{Res}(Q_n,Q_{n-1}),$$

since Q_n is of degree n. By induction we deduce,

$$\mathrm{Res}(Q_{n+1}, Q_n) = (-1)^{\frac{n(n+1)}{2}} \prod_{i=1}^{n} a_i^{2i} \mathrm{Res}(Q_1, Q_0).$$

As $Q_0 = 1$, we get

$$\mathrm{Res}(Q_{n+1}, Q_n) = (-1)^{\frac{n(n+1)}{2}} \prod_{i=1}^{n} a_i^{2i}.$$

As the zeros of Q_n are the eigenvalues of $J^{(n)}$, we get the claim. □

From (6.12) and (6.14), we deduce

$$\frac{1}{n^2} \log \mathbb{P}\big(||J_n - K_n||_\infty < \delta/2\big) \geq -\tilde{I}_\beta(\nu_n, \nu_{n-1}) - \varepsilon_n(\delta),$$

where $\tilde{I}_\beta$ is defined by

$$\forall \nu, \eta \in \mathcal{P}(\mathbb{R}),\ \tilde{I}_\beta(\nu, \eta) = \frac{\beta}{4}\nu(x^2) - \frac{\beta}{2}\int_{x \neq y} \log|x-y| d\nu(x) d\eta(y) - c_\beta. \tag{6.15}$$

Since $\nu_n = n^{-1}\sum_{i=1}^{n} \delta_{z_i^{(n)}}$, where $z_i^{(n)}$ are the $1/(n+1)$-quantiles of the target measure μ, we get by a similar argument as in [4],

$$\lim_{n \to +\infty} \tilde{I}_\beta(\nu_n, \nu_{n-1}) = I_\beta(\mu).$$

This ends the proof of the lower bound in view of (6.10). □

We now turn to the proof of the upper bound of Theorem 6.1.1.

Proof of the upper bound. As discussed in the introduction, we will prove the upper bound by a contraction argument. Recall that $\mathcal{P}_{\mathrm{stat}}$ the set of stationary probability measures ρ on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$ endowed with the Borel σ-algebra of the product topology.

We define $\hat{\mathcal{P}}_{\mathrm{stat}}$ the set of probability measures π on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}} \times [0,1]$, endowed with the Borel σ-algebra of the product topology, such that the marginal on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$ is in $\mathcal{P}_{\mathrm{stat}}$. We will see an element π in $\hat{\mathcal{P}}_{\mathrm{stat}}$, as the law of an infinite Jacobi matrix, $J = (v_k, w_k)_{k \in \mathbb{Z}}$, with stationary coefficients, together with a random variable $T \in [0,1]$, which we will call the "time".

For a finite Jacobi matrix K_n with coefficients $(b_k, a_k)_{1 \leq k \leq n}$, where $a_n = 0$ by convention, we associate the law of $((b_{k+T[n]}, a_{k+T[n]})_{k \in \mathbb{Z}}, T/n)$, with T uniformly sampled in $\{1, ..., n\}$. This defines a probability measure in $\hat{\mathcal{P}}_{\mathrm{stat}}$ which we denote by π_{K_n}. For the Jacobi matrix J_n given by the tridiagonal representation of the β-ensemble with quadratic potential in (6.1), we denote by π_n the random probability measure π_{J_n}.

Recall the Lévy-metric d_L on probability measures on $\mathbb{R}^n$,

$$d_L(\mu, \nu) = \inf\{\delta > 0 : \forall F \text{ closed},\ \mu(F^\delta) \leq \nu(F) + \delta\},$$

where F^δ denote the δ-neighborhood, which we take here with respect to the sup norm $||\ ||_\infty$. This metric induces the weak topology on the set of probability measures on $\mathbb{R}^n$. Thus, the weak topology on $\hat{\mathcal{P}}_{\text{stat}}$ with respect to the product topology is metrizable by the metric $\hat{d}$ defined by,

$$\hat{d}(\pi, \pi') = \frac{1}{1+s},$$

where s is the supremum over the radii $r > 0$ such that

$$d_L(\pi \circ p_r^{-1}, \pi' \circ p_r^{-1}) < 1/r,$$

with p_r the projection $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}^{\mathbb{Z}} \times [0,1] \to \mathbb{R}^{\mathbb{Z}_r} \times \mathbb{R}^{\mathbb{Z}_r} \times [0,1]$, where $\mathbb{Z}_r = \{k \in \mathbb{Z} : |k| \leq r\}$.

We will prove the following large deviations upper bound.

6.4.3 Lemma. *Let $\hat{\mathcal{P}}$ be the subset of $\hat{\mathcal{P}}_{stat}$ of probability measures such that their marginal on $[0,1]$ is the uniform measure. For any closed subset F of $\hat{\mathcal{P}}_{stat}$,*

$$\limsup_{n\to+\infty} \frac{1}{n^2} \log \mathbb{P}(\pi_{J_{n,\beta}} \in F) \leq - \inf_{\pi \in F \cap \hat{\mathcal{P}}} I_\beta(\mu_{\rho(\pi)}),$$

where $\rho(\pi)$ denotes the marginal of π on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$, and I_β is as in Theorem 6.1.1.

Before going into the proof, we will show the following density result.

6.4.4 Lemma. *For any $\pi \in \hat{\mathcal{P}}$, there is a sequence of Jacobi matrices $(K_n)_{n\in\mathbb{N}}$, such that $(\pi_{K_n})_{n\in\mathbb{N}}$ converges weakly to π.*

Proof. Let $\pi \in \hat{\mathcal{P}}_{\text{stat}}$ be the law of (J, T) on some probability space. Let $p : [0,1] \times \mathcal{B} \to [0,1]$ be the conditional kernel of J given T, where $\mathcal{B}$ denotes the Borel σ-algebra of $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$.

We can write $\pi = \mathcal{U}.p$, where $\mathcal{U}$ denotes the uniform measure on $[0,1]$, and $\mathcal{U}.p$ is the probability measure on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}} \times [0,1]$, defined for any Borel subsets A and B of respectively $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$ and $[0,1]$, by

$$\mathcal{U}.p(A \times B) = \int_B p(t, A) d\mathcal{U}(t).$$

At the price of regularizing p, by setting for any $(t, A) \in [0,1] \times \mathcal{B}$,

$$p_\varepsilon(t, A) = \frac{1}{2\varepsilon} \int_{|t-s|<\varepsilon} p(s, A) ds,$$

and replacing π by $\pi_\varepsilon = \mathcal{U}.p_\varepsilon$, we can assume that for any bounded continuous function f on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}} \times [0,1]$,

$$t \in [0,1] \mapsto \int f(x,t) p(t, dx),$$

is continuous. Let $n, d \in \mathbb{N}$, $d \leq n$. Set $N = \lfloor n/d \rfloor$. For any $k \in \{0, ..., N\}$, let $x^{(k)} \in \mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$ be independent random variables with law $p(k/n, .)$. For any $j \in \{1, ..., n\}$, define

$$x_j = \begin{cases} x_j^{(\lfloor j/d \rfloor)} & \text{if } j \leq Nd, \\ (0,1) & \text{if } j > Nd, \end{cases}$$

and with the second coordinate of x_n set to be 0. Let K_n be the Jacobi matrix with coefficients $(x_j)_{1\leq j\leq n}$, that is $(K_n(j,j),K_n(j,j+1)) = x_j$ for any $j \in \{1,...,n\}$. We will prove that if d is chosen so that $d \to +\infty$, and $n/d \to +\infty$, then

$$\pi_{K_n} \underset{n\to+\infty}{\rightsquigarrow} \pi,$$

weakly, almost surely. Indeed, let $r > 0$ and $f : \mathbb{R}^{\mathbb{Z}_r} \times \mathbb{R}_+^{\mathbb{Z}_r} \times [0,1] \to \mathbb{R}$ be a continuous function bounded by 1 and vanishing at ∞. In particular, f is uniformly continuous. Neglecting the border terms, and using the regularity of f, we see that

$$\pi_{J_n}(f) = \frac{1}{n}\sum_{i=1}^{n} f\big((x_{i+j[n]})_{|j|\leq r}, i/n\big) = \frac{1}{n}\sum_{k=1}^{N}\sum_{\lfloor i/d\rfloor = k} f\big((x^{(k)}_{i+j[n]})_{|j|\leq r}, kd/n\big) + O\Big(\frac{r}{d}\Big).$$

Denote by $F(x^{(1)},...,x^{(N)})$ the first term on the right-hand side of the above equation. Note that as f is bounded by 1,

$$F(Y) - F(X) \leq \frac{2d}{n},$$

if X and Y differ only by one coordinate. By the bounded difference inequality (see [76, Theorem 6.2]), we deduce

$$\mathbb{P}\big(|F(x^{(1)},...,x^{(N)}) - \mathbb{E}F(x^{(1)},...,x^{(N)})| > t\big) \leq 2e^{-c\frac{t^2 n}{d}}, \tag{6.16}$$

for some constant $c > 0$. But, by the stationarity of π we have

$$\mathbb{E}F(x^{(1)},...,x^{(N)}) = \frac{d-1}{n}\sum_{k=1}^{N}\int f(x,kd/n)p(kd/n,dx).$$

As $t \mapsto \int f(x,t)p(t,dx)$, is continuous by deduce that for $d/n \to +\infty$,

$$\mathbb{E}F(x^{(1)},...,x^{(N)}) \underset{n\to+\infty}{\longrightarrow} \pi(f).$$

From (6.16) and Borel-Cantellli-Lemma, we deduce that if d and $n/d \to +\infty$, $\pi_{K_n}(f)$ converges almost surely to $\pi(f)$. By choosing a dense countable subset of the space of continuous function vanishing at infinity $\mathbb{R}^{\mathbb{Z}_r} \times \mathbb{R}_+^{\mathbb{Z}_r} \times [0,1] \to \mathbb{R}$, we deduce that almost surely $\pi_{K_n} \circ p_r^{-1}$ converges vaguely to $\pi \circ p_r^{-1}$, and hence weakly as $\pi \circ p_r^{-1}$ is a probability measure. Therefore almost surely π_{K_n} converges weakly to π

□

We are now ready to prove Lemma 6.4.3.

Proof of Lemma 6.4.3. The sequence $(\pi_n)_{n\in\mathbb{N}}$ is exponentially tight at the scale n^2, since one can find a $\tau > 0$ such that

$$\limsup_{n\to+\infty}\frac{1}{n^2}\log\mathbb{E}e^{\tau n \mathrm{tr} J_n^2} < +\infty.$$

This entails also, by the lower semi-continuity of $\pi \mapsto \mathbb{E}_\pi(2w_o^2 + v_o^2)$, that the upper bound holds for any $\pi \in \hat{\mathcal{P}}_{\mathrm{stat}}$ such that $\mathbb{E}_\pi(2w_o^2 + v_o^2) = +\infty$. Moreover, as the marginal of π_n on $[0,1]$ is deterministic and converges weakly to the uniform measure on $[0,1]$, we see that the upper bound of Lemma 6.4.3 holds for any $\pi \notin \hat{\mathcal{P}}$.

Let now $\delta > 0$ and $\pi \in \hat{\mathcal{P}}$ such that $\mathbb{E}_\pi(2w_o^2 + v_o^2) < +\infty$. By Lemma 6.4.4, there is a sequence of Jacobi matrices K_n such that π_{K_n} converges weakly to π. By Lemma 6.2.1, we know that ν_n, the empirical spectral measure of J_n, converges weakly to μ_ρ, where ρ is the marginal of π on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$. For n large enough we have

$$\mathbb{P}\big(\hat{d}(\pi_n, \pi) < \delta\big) \leq \mathbb{P}\big(\hat{d}(\pi_n, \pi_{K_n}) < 2\delta\big).$$

Thus, there is a $r > 1/2\delta - 1$ such that

$$\mathbb{P}\big(\hat{d}(\pi_n, \pi) < \delta\big) \leq \mathbb{P}\big(d_L(\pi_n \circ p_r^{-1}, \pi_{K_n} \circ p_r^{-1}) < 1/r\big). \tag{6.17}$$

Let

$$F = \bigcup_{l=0}^{n-1} \big\{(x,y,t) : ||((x_k,y_k)_{|k|\leq r}, t) - ((X_{k+l[n]}, Y_{k+l[n]})_{|k|\leq r}, l/n)||_\infty \leq 1/r\big\}.$$

As $\pi_n(F) = 1$, we have

$$\mathbb{P}\big(d_L(\pi_n \circ p_r^{-1}, \pi_{K_n} \circ p_r^{-1}) < 1/r\big) \leq \mathbb{P}\big(\pi_{K_n}(F^{1/r}) \geq 1 - 1/r\big). \tag{6.18}$$

Performing a union bound we get,

$$\mathbb{P}\big(\pi_n(F^{1/r}) \geq 1 - 1/r\big) \leq \sum_{I,\sigma} \mathbb{P}(E_{I,\sigma}), \tag{6.19}$$

where the sum is over all subsets I of $\{1,...,n\}$ such that $|I| \geq n(1 - r^{-1})$, and functions $\sigma : I \to \{1,...,n\}$ such that $|\sigma(i) - i| \leq 2n/r$, and where $E_{I,\sigma}$ is the event,

$$E_{I,\sigma} = \Big\{\forall i \in I,\ \big|\big|(b_i, a_i) - (X_{\sigma(i)}, Y_{\sigma(i)})\big|\big|_\infty < \frac{2}{r}\Big\}.$$

But, by Lemma 6.4.1,

$$\mathbb{P}(E_{I,\sigma}) \leq e^{-n^2(H_n - c^I_{\beta,n})},$$

where

$$H_n = \frac{\beta}{4n}(1-4\delta)\Big(\sum_{i\in I} b_i^2 + 2\sum_{i \in I}(a_i+\delta)^2\Big) - \frac{\beta}{n}\sum_{i\in I}\frac{\sigma(i)}{n}\log(a_i + \delta) - \varepsilon_n(\delta),$$

with $\varepsilon_n(\delta)$ is some function, which vary along the proof, such that $\limsup_n \varepsilon_n(\delta) = o(1)$, and

$$c^I_{\beta,n} = \frac{1}{n^2}\log\prod_{i\in I}\frac{(\beta n/2)^{\frac{i\beta}{2}}}{\Gamma(i\beta/2)},$$

But one can see that

$$c^I_{\beta,n} = c_{\beta,n} + O(r^{-1}),$$

uniformly in n and $I \subset \{1,...,n\}$ such that $|I| \geq n(1-r^{-1})$. Thus by (6.13),

$$\mathbb{P}(E_{I,\sigma}) \leq e^{-n^2(H_n - c_\beta)}, \tag{6.20}$$

Bounding roughly the number of terms in the sum on the right-hand side of (6.19) by $2^n n^n$, we get by (6.20), (6.17) and (6.18),

$$\mathbb{P}\big(\hat{d}(\pi_n,\pi)<\delta\big)\le e^{-n^2(H_n-c_\beta)}. \tag{6.21}$$

Let $K_n^{I,\delta}$ be the Jacobi matrix with diagonal coefficients $(b_k\mathbb{1}_{k\in I})_{1\le k\le n}$ and up-diagonal coefficients $(a_k+\delta)$ if $k\in I$, and 1 otherwise. Let $\nu_n^{I,\delta}$ be its empirical spectral measure. We have,

$$\frac{1}{n}\Big(\sum_{i\in I} b_i^2+2\sum_{i\in I}(a_i+\delta)^2\Big)+\frac{2|I^c|}{n}=\nu_n^{I,\delta}(x^2).$$

Besides,

$$\frac{1}{n}\sum_{i\in I}\frac{\sigma(i)}{n}\log(a_i+\delta)\le\frac{1}{n}\sum_{i\in I}\frac{i}{n}\log(a_i+\delta)+\frac{2}{rn}\sum_{i\in I}\log(1+a_i+\delta).$$

By Lemma 6.4.2, we recognize,

$$\frac{1}{n}\sum_{i\in I}\frac{i}{n}\log(a_i+\delta)=\frac{(n-1)}{2n}\int\log|x-y|d\nu_n^{I,\delta}(x)d\nu_{n-1}^{I,\delta}(y).$$

Using the bound $\log(1+x)\le x^2+1$,

$$\frac{2}{rn}\sum_{i\in I}\log(1+a_i+\delta)\le\frac{1}{r}(\nu_n^{I,\delta}(x^2)+2).$$

Therefore, we can write for n large enough,

$$H_n\ge(1-O(r^{-1}))\frac{\beta}{4}\nu_n^{I,\delta}(x^2)-\frac{\beta}{2}\int\log|x-y|d\nu_n^{I,\delta}(x)d\nu_{n-1}^{I,\delta}(y)-\varepsilon_n(\delta).$$

As $\log|x-y|\le x^2+y^2+2$, we have

$$H_n\ge(1-O(r^{-1}))\int f(x,y)d\nu_n^{I,\delta}(x)d\nu_{n-1}^{I,\delta}(y)-\varepsilon_n(\delta),$$

with

$$f(x,y)=\frac{\beta}{2}\Big(\frac{1}{4}x^2+\frac{1}{4}y^2\Big)-\frac{\beta}{2}\log|x-y|.$$

Fix $M>1$. Then,

$$H_n\ge(1-O(r^{-1})\int(f(x,y)\wedge M)d\nu_n^{I,\delta}(x)d\nu_{n-1}^{I,\delta}(y)-\varepsilon_n(\delta). \tag{6.22}$$

As $f\wedge M$ is bounded and continuous,

$$F_M:(\mu,\nu)\mapsto\int(f(x,y)\wedge M)d\mu(x)d\nu(y),$$

is continuous for the weak topology. By (6.21), we deduce,

$$\frac{1}{n^2}\log\mathbb{P}\big(\hat{d}(\pi_n,\pi)<\delta\big)\le-(1-O(r^{-1}))F_M(\nu_n^{I,\delta},\nu_{n-1}^{I,\delta})+c_\beta+\varepsilon_n(\delta).$$

We claim that

$$\lim_{\delta\to 0}\limsup_{n\to+\infty} d(\mu,\nu_n^{I,\delta}) = 0.$$

On one hand, by the interlacing inequality (1.3),

$$d_{KS}(\nu_n^{I,0},\nu_n) \le \frac{1}{n}\mathrm{rank}(K_n - K_n^I),$$

where d_{KS} denote the Kolmogorov-Smirnov distance. Since the rank of a matrix is upper bounded by the number of non-zeros entries, we get

$$d_{KS}(\nu_n,\nu_{n,0}^I) \le \frac{|I^c|}{n} \le \frac{1}{r}.$$

On the other hand, by Hoeffman-Wielandt inequality (1.2),

$$\mathcal{W}_2(\nu_n^{I,\delta},\nu_n^{I,0}) \le \delta^{1/2},$$

where $\mathcal{W}_2$ denotes the L^2-Wasserstein distance. As ν_n converges weakly to μ, and $r > 1/2\delta - 1$, we get by the triangular inequality,

$$\lim_{\delta\to 0}\limsup_{n\to+\infty} d(\mu,\nu_n^{I,\delta}) = 0.$$

Therefore, from the lower semi-continuity of F_M, we deduce that there is a function $h(r^{-1})$ such that $h(r^{-1}) \to 0$ as $r \to +\infty$, so that,

$$\limsup_{n\to+\infty}\frac{1}{n^2}\log\mathbb{P}\big(\hat{d}(\pi_n,\pi)<\delta\big) \le -(1-O(r^{-1}))F_M(\mu,\mu) + h(r^{-1}).$$

Taking the $\limsup$ as δ goes to 0 (and hence r to $+\infty$), and then M to $+\infty$, we get by monotonous convergence,

$$\lim_{\delta\to 0}\limsup_{n\to+\infty}\frac{1}{n^2}\log\mathbb{P}\big(\hat{d}(\pi_n,\pi)<\delta\big) \le -I_\beta(\mu).$$

. $\square$

We come back now to the proof of the upper bound of Theorem 6.1.1. As $\rho\in\mathcal{E}_2\mapsto\mu_\rho$ is continuous by Lemma 6.2.1, and

$$\{\pi\in\mathcal{P}_{\text{stat}} : \rho(\pi)\in\mathcal{E}_2\},$$

contains the domain of the rate function upper bound of π_n by Lemma 6.4.3, we deduce by the contraction principle (see [43, Theorem 4.2.1], remark (c)) that we have a LDP upper bound for $(\mu_n)_{n\in\mathbb{N}}$, with speed n^2 and rate function

$$H_\beta(\mu) = \inf\{I_\beta(\mu_{\rho(\pi)}) : \mu_{\rho(\pi)} = \mu, \pi\in\hat{\mathcal{P}}\} = \inf\{I_\beta(\mu) : \mu_{\rho(\pi)} = \mu, \pi\in\hat{\mathcal{P}}\},$$

where for $\pi\in\hat{\mathcal{P}}_{\text{stat}}$, we denote by $\pi(\rho)$ its marginal on $\mathbb{R}^{\mathbb{Z}}\times\mathbb{R}_+^{\mathbb{Z}}$. To prove that $H_\beta = I_\beta$, we need to show that for any $\mu\in\mathcal{P}(\mathbb{R})$ such that $\mu(x^2)<+\infty$, there is some $\pi\in\hat{\mathcal{P}}$ such that $\mu_{\rho(\pi)} = \mu$. This will indeed be sufficient since if $\mu(x^2) = +\infty$, both $I_\beta(\mu)$ and $H_\beta(\mu)$ are infinite. Thus, to complete the proof of the upper bound of Theorem 6.2.1, we only need to prove the following lemma.

$\square$

6.4.5 Lemma. *Let $\mu \in \mathcal{P}(\mathbb{R})$ such that $\mu(x^2) < +\infty$. There is an admissible measure $\pi \in \hat{\mathcal{P}}$, such that $\mu = \mu_\rho$, where ρ is the marginal of π on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}}$.*

Proof. Let $\mu_\varepsilon = \mu * \nu_\varepsilon$, where ν_ε is the law of a Gaussian random variable with variance ε^2. Then, $\mu_\varepsilon(x^2) = \mu(x^2) + \varepsilon^2$. Just as in the proof of the lower bound of Theorem 6.1.1, let $n \in \mathbb{N}$ and define for any $k \in \{1, ..., n\}$,

$$z_k^{(n)} = F_\varepsilon^{(-1)}(\frac{k}{n+1}),$$

with $F_\varepsilon^{(-1)}$ the inverse of the distribution function of μ_ε. Set

$$\nu_n = n^{-1} \sum_{k=1}^n \delta_{z_k^{(n)}}.$$

As μ_ε has no atoms the measure ν_n and ν_{n-1} have atoms which interlace strictly. The Geronimus-Wendroff theorem (see [84, Chapter 1, 1.2 §6]) tells us that there is a Jacobi matrix $K_n = (b_k, a_k)_{1 \le k \le n}$, with the convention $a_n = 0$, such that ν_n and ν_{n-1} are the spectral measures of J_n and $J_n^{(n-1)}$, the $n-1$ principal matrix of J_n, respectively. Let T be a random variable uniformly sampled in $\{1, ..., n\}$.

$$\mathbb{E}|b_{k+T[n]}|^2 + 2\mathbb{E}|a_{k+T[n]}|^2 \le \nu_n(x^2) \le \frac{n+1}{n}\mu_\varepsilon(x^2). \tag{6.23}$$

We deduce that the law of $(b_{k+T[n]}, a_{k+T[n]}, T/n)_{k\in\mathbb{Z}}$ is tight for the product topology. Thus, by Prokhorov's theorem, there is a subsequence which converges to some π_ε. As ν_n converges weakly to μ_ε by construction, we have by continuity of the spectral measure (see Lemma 6.2.1) $\mu_{\rho_\varepsilon} = \mu_\varepsilon$. The lower semi-continuity of $\pi \mapsto \mathbb{E}_\pi(v_k^2 + 2w_k)$ yields, taking the lim inf in (6.23),

$$\mathbb{E}_{\pi_\varepsilon}(v_k^2 + 2w_k^2) \le \mu_\varepsilon(x^2).$$

But as $\mu_\varepsilon(x^2)$ is bounded in ε, $(\rho_\varepsilon)_{\varepsilon>0}$ is tight as probability measures on $\mathbb{R}^{\mathbb{Z}} \times \mathbb{R}_+^{\mathbb{Z}} \times [0,1]$. Thus, using the fact that μ_ε converges to μ as $\varepsilon \to 0$, and using again the continuity of the spectral measure for the weak topology, we deduce that there is a measure in $\hat{\mathcal{P}}$ whose spectral measure is μ. $\square$

6.4.6 Remark. One can wonder if the lower bound LDP for $(\pi_n)_{n\in\mathbb{N}}$ holds with the rate function of Lemma 6.4.3. We actually do not how to derive such a lower bound, because it implies to know how to control, given a measure π, and an approximating sequence π_{K_n}, the eigenvalue spacings of K_n, in order to get the uniform integrability of the function $(x,y) \mapsto \log(|x-y| \wedge 1)$ under $\mu_{K_n} \otimes \mu_{K_{n-1}}$.

6.5 LDP for the discrete Schrödinger operator

We will prove in this section the variational formula of Proposition 6.3.1. We will obtain it by deriving a second LDP for $(\pi_n)_{n\in\mathbb{N}}$, following the lines of the proof of Lemma 6.4.3, by keeping in the estimates of the probability of deviations of π_n, a formulation in terms of the coefficients and not of the spectral measure. We prove the following result.

6.5.1 Theorem. *We recall that we denote by $\hat{\mathcal{P}}$ the probability measures of $\hat{\mathcal{P}}_{stat}$ whose marginal on $[0,1]$ is the uniform law. $(\pi_n)_{n\in\mathbb{N}}$ follows a LDP with speed n^2 with respect to the weak topology of $\hat{\mathcal{P}}_{stat}$, and good rate function $\mathcal{I}_\beta$ defined for any $\pi \in \hat{\mathcal{P}}$ such that $\mathbb{E}_\pi(2w_o^2+v_o^2) < +\infty$, by*

$$\mathcal{I}_\beta(\pi) = \mathbb{E}_\pi(2w_o^2+v_o^2) - \beta\mathbb{E}_\pi T\log w_o - c_\beta,$$

where under $\mathbb{P}_\pi$, $((w_k,v_k)_{k\in\mathbb{Z}},T)$ has law π, and by $\mathcal{I}_\beta(\pi)=+\infty$ otherwise

Proof. **Upper bound:** With the same argument as in the proof of Lemma 6.4.3, we can restrict ourself to derive the upper bound for deviations around measures $\pi\in\hat{\mathcal{P}}$ such that $\mathbb{E}_\pi(2w_o^2+v_o^2)<+\infty$. Let π be such a measure and let $\delta>0$. We showed in the proof of Lemma 6.4.3 that for some $r>1/2\delta-1$,

$$\mathbb{P}\big(\hat{d}(\pi_n,\pi)<\delta\big) \le \sum_{I,\sigma} p_{E,I}.$$

where the sum is over all subsets I of $\{1,...,n\}$ such that $|I|\ge n(1-r^{-1})$, and functions $\sigma: I\to\{1,...,n\}$ such that $|\sigma(i)-i|\le 2n/r$, and

$$p_{E,I} \le e^{-n^2(H_n-c_\beta)},$$

with

$$H_n = \frac{1}{n}(1-O(\delta))\big(\sum_{i\in I} b_i^2 + 2\sum_{i\in I}(a_i+\delta)^2\big) - \frac{\beta}{n}\sum_{i\in I}\frac{i}{n}\log a_i - \varepsilon_n(\delta),$$

where $\varepsilon_n(\delta)$ is some function which will vary along the proof such that $\limsup_n \varepsilon_n(\delta)=o(1)$. Define as in Lemma 6.4.3, the Jacobi matrix $K_n^{I,\delta}$, with diagonal coefficients $b_k\mathbb{1}_{k\in I}$ and up-diagonal coefficients $a_k+\delta$ if $k\in I$, and 1 otherwise, and $\pi_n^{I,\delta}$ the measure in $\hat{\mathcal{P}}_{\text{stat}}$ associated. We recognize,

$$\frac{1}{n}\big(\sum_{i\in I} b_i^2 + 2\sum_{i\in I}(a_i+\delta)^2\big) + \frac{2|I^c|}{n} = \mathbb{E}_{\pi_{n,\delta}^I}(2w_o^2+v_o^2),$$

and

$$\frac{1}{n}\sum_{i\in I}\frac{i}{n}\log a_i = \mathbb{E}_{\pi_{n,\delta}^I} T\log w_o,$$

Thus,

$$H_n \ge \big(1+O(r^{-1})\big)\mathbb{E}_{\pi_{n,\delta}^I}(2w_o^2+v_o^2) - \beta\mathbb{E}_{\pi_{n,\delta}^I}T\log w_o - \varepsilon_n(\delta).$$

If we define

$$\varphi((w_k,v_k)_{k\in\mathbb{Z}},t) = (2w_o^2+v_o^2) - \frac{\beta}{2}t\log w_o,$$

we have the bound,

$$H_n \ge \big(1+O(r^{-1})\big)\int \varphi d\pi_{n,\delta}^I - \varepsilon_n(\delta).$$

We claim that

$$\lim_{r\to+\infty}\limsup_{n\to+\infty} d(\pi_{n,\delta}^I,\pi_{K_n}) = 0,$$

Indeed,

$$\mathcal{W}_2(\pi^I_{n,\delta} \circ p_r^{-1}, \pi_{K_n} \circ p_r^{-1}) \leq ||K_n^{I,\delta} - K_n||_\infty \leq \delta,$$

where $\mathcal{W}_2$ denotes the L^2-Wasserstein distance on probability measures on $\mathbb{R}^{\mathbb{Z}_r} \times \mathbb{R}_+^{\mathbb{Z}_r} \times [0,1]$, with respect to the sup norm. Thus, the same argument as in the proof of Lemma 6.4.3 yields, as π_{K_n} converges to π,

$$\lim_{\delta \to 0} \limsup_{n \to +\infty} \frac{1}{n^2} \log \mathbb{P}\big(d(\pi_n, \pi) < 2\delta\big) \leq -\int \varphi d\pi + c_\beta,$$

which gives the upper bound.

Lower bound: Let $\pi \in \hat{\mathcal{P}}$ and such that $\mathcal{I}_\beta(\pi) < +\infty$. At the price of truncating the weights, we can assume that under π, $|v_k| \leq R$, and $R^{-1} \leq w_k \leq R$, for any $k \in \mathbb{Z}$ and for some $R > 0$. There is a sequence K_n of Jacobi matrices such that π is the weak limit of π_{K_n}. Once again, at the price of truncating the coefficients, we can assume that $|b_k| \leq R$ and $R^{-1} \leq a_k \leq R$, for any k.

Let $\delta > 0$. For n large enough,

$$\mathbb{P}(d(\pi_n, \pi) < 2\delta) \geq \mathbb{P}(d(\pi_n, \pi_{K_n}) < \delta).$$

But one can find a function η with $\eta(\delta) \to 0$ when $\delta \to 0$, such that

$$\mathbb{P}(d(\pi_n, \pi_{K_n}) < \delta) \geq \mathbb{P}\big(||J_n - K_n||_\infty < \eta(\delta)\big).$$

By Lemma 6.4.1 we deduce,

$$\mathbb{P}\big(||J_n - K_n||_\infty < \eta(\delta)\big) \geq e^{-n^2 I_n - \varepsilon_n(\delta)},$$

where

$$I_n = \frac{1}{2n} \sum_{i=1}^{n} (b_i^2 + 2a_i^2) - \beta \sum_{i=1}^{n-1} \frac{i}{n} \log a_i - c_\beta,$$

which we can re-write as,

$$I_n = \frac{1}{2} \mathbb{E}_{\pi_{J_n}}(v_o^2 + 2w_o^2) - \beta \mathbb{E}_{\pi_{J_n}} T \log w_o - c_\beta.$$

As under π, as well as π_{K_n}, we have $|v_o| \leq R$, and $R^{-1} \leq a_k \leq R$, and π_{K_n} converges weakly to π, we get by dominated convergence,

$$\lim_{\delta \to 0} \liminf_{n \to +\infty} \frac{1}{n^2} \log \mathbb{P}(||J_n - K_n||_\infty < \eta(\delta)) \geq -\frac{1}{2} \mathbb{E}_\pi(v_o^2 + 2w_o^2) - \beta \mathbb{E}_\pi T \log w_o - c_\beta.$$

□

We can now give a proof of Proposition 6.3.1.

Proof of Proposition 6.3.1. As the map $\pi \mapsto \mu_{\rho(\pi)}$ is continuous by Lemma 6.2.1, we get by the contraction principle, a LDP for the spectral measure of J_n. Comparing the rate function with the rate function of Theorem 6.1.1, we get the variational formula of Proposition 6.3.1. □

6.6 Conclusion and perspectives

We showed the one can take advantage of the tridiagonal representation of the classical Gaussian ensembles to obtain by contraction the LDP for the spectral measure. This approach can be contemplated for any models of random matrix conjugated to a sparse matrix with a tractable law. This was actually the approach adopted by Bordenave and Caputo [29] for the LDP of the spectral measure of Wigner matrices without Gaussian tails, where a LDP for the spectral measure of a sparse matrix was needed. Moreover, in the same spirit as the LDP we gave for the β-ensembles with a quadratic potential, one can consider other classical matrix models which have representation as a sparse matrix with an explicit law, like the Laguerre ensemble, or Haar matrices and their pentadiagonal representation (see [64, Theorem 1.2]). We believe this approach could also be extended to unitarily invariant matrix models with a general potential V. Indeed, these models gives rise to a tridiagonal representation but with dependent coefficients.

In the same spirit as the proof we gave of the LDP of the spectral measure of GUE and GOE, one can wonder if a similar "objective method" can be carried out in the large deviations analysis of the edges of the spectrum of GUE and GOE matrices, which would recover the LDP known for the largest eigenvalue of β-ensembles by [16]. The framework developed by Ramirez, Rider, and Virag in [83] who considered the scaling limit at the edge of the tridiagonal representation of β-ensembles, may be a good starting point for an objective method in the large deviations of the edges of GUE and GOE.

Bibliography

[1] M. Aizenman, R. Sims, and S. Warzel. Stability of the absolutely continuous spectrum of random Schrödinger operators on tree graphs. *Probab. Theory Related Fields*, 136(3):363–394, 2006.

[2] D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.

[3] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[4] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probabilty theory and related fields*, 108:517–542, 1997.

[5] E. Artin. *The gamma function*. Translated by Michael Butler. Athena Series: Selected Topics in Mathematics. Holt, Rinehart and Winston, New York-Toronto-London, 1964.

[6] F. Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.*, 21:Paper No. 32, 49, 2016.

[7] F. Augeri. On the large deviations of traces of random matrices. *arXiv:1605.03894*, May 2016.

[8] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[9] Z. D. Bai and J. Yao. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, 11(6):1059–1092, 2005.

[10] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16(4):1729–1741, 1988.

[11] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[12] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[13] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[14] F. Barthe, P. Cattiaux, and C. Roberto. Concentration for independent random variables with heavy tails. *AMRX Appl. Math. Res. Express*, (2):39–60, 2005.

[15] S. T. Belinschi, Hari. Bercovici, and M. Février. Outliers in the spectrum of large deformed unitarily invariant models. *to appear in the Annals of Probability*, arXiv:1412.4916.

[16] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001.

[17] G. Ben Arous and A. Guionnet. The spectrum of heavy tailed random matrices. *Comm. Math. Phys.*, 278(3):715–751, 2008.

[18] F. Benaych-Georges, A. Guionnet, and M. Maida. Large deviations of the extreme eigenvalues of random deformations of matrices. *Probab. Theory Related Fields*, 154(3-4):703–751, 2012.

[19] F. Benaych-Georges, A. Guionnet, and C. Male. Central limit theorems for linear statistics of heavy tailed random matrices. *Comm. Math. Phys.*, 329(2):641–686, 2014.

[20] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank pertubation of large random matrices. *Advances in Mathematics*, 227:494–521, 2011.

[21] H. Bercovici and D. Voiculescu. Free convolution of measures with unbounded support. *Indiana Univ. Math. J.*, 42(3):733–773, 1993.

[22] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[23] P. Biane. On the free convolution with a semi-circular distribution. *Indiana Univ. Math. J.*, 46(3):705–718, 1997.

[24] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[25] S. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields*, 107(3):383–400, 1997.

[26] S. G. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab.*, 27(4):1903–1921, 1999.

[27] S. G. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[28] C. Bordenave. *Spectrum of random graphs.* https://www.math.univ-toulouse.fr/ bordenave/coursSRG.pdf.

[29] C. Bordenave and P. Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.*, 42(6):2454–2496, 2014.

[30] C. Bordenave and P. Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probab. Theory Related Fields*, 163(1-2):149–222, 2015.

[31] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of large random reversible Markov chains: heavy-tailed weights on the complete graph. *Ann. Probab.*, 39(4):1544–1590, 2011.

[32] C. Borell. Tail probabilities in Gauss space. In *Vector space measures and applications (Proc. Conf., Univ. Dublin, Dublin, 1977), I*, volume 644 of *Lecture Notes in Math.*, pages 73–82. Springer, Berlin-New York, 1978.

[33] C. Borell. On polynomial chaos and integrability. *Probab. Math. Statist.*, 3(2):191–203, 1984.

[34] G. Borot and A. Guionnet. Asymptotic expansion of β matrix models in the one-cut regime. *Comm. Math. Phys.*, 317(2):447–483, 2013.

[35] M. Capitaine, C. Donati-Martin, D. Féral, and M. Février. Free convolution with a semicircular distribution and eigenvalues of spiked deformations of Wigner matrices. *Electron. J. Probab.*, 16:no. 64, 1750–1792, 2011.

[36] R. Carmona and J. Lacroix. *Spectral theory of random Schrödinger operators.* Probability and its Applications. Birkhäuser Boston, Inc., Boston, MA, 1990.

[37] D. Chafaï, A. Hardy, and M. Maïda. Concentration for Coulomb gases and Coulomb transport inequalities. *arXiv:1610.00980*, November 2016.

[38] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[39] S. Chatterjee and S. R. S. Varadhan. Large deviations for random matrices. *Commun. Stoch. Anal.*, 6(1):1–13, 2012.

[40] F. Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264 of *Graduate Texts in Mathematics*. Springer, London, 2013.

[41] P. A. Deift. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3 of *Courant Lecture Notes in Mathematics*. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.

[42] J. Delon, J. Salomon, and A. Sobolevski. Local matching indicators for transport problems with concave costs. *SIAM J. Discrete Math.*, 26(2):801–827, 2012.

[43] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[44] R. Dobrushin, P. Groeneboom, and M. Ledoux. *Lectures on probability theory and statistics*, volume 1648 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, Edited by P. Bernard.

[45] H. Döring and P. Eichelsbacher. Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices. *Electron. J. Probab.*, 14:no. 92, 2636–2656, 2009.

[46] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.

[47] F. J. Dyson. The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics. *J. Mathematical Phys.*, 3:1199–1215, 1962.

[48] D. E. Edmunds and H. Triebel. *Function spaces, entropy numbers, differential operators*, volume 120 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[49] P. Eichelsbacher and J. Sommerauer. Moderate deviations for traces of words in a multi-matrix model. *Electron. Commun. Probab.*, 14:572–586, 2009.

[50] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[51] F. Gamboa, J. Nagel, and A. Rouault. Sum rules and large deviations for spectral measures on the unit circle. *to appear in Random Matrices: Theory and Applications*, arXiv:1601.08135.

[52] F. Gamboa, J. Nagel, and A. Rouault. Sum rules via large deviations. *J. Funct. Anal.*, 270(2):509–559, 2016.

[53] N. Gantert, K. Ramanan, and F. Rembart. Large deviations for weighted sums of stretched exponential random variables. *Electron. Commun. Probab.*, 19:no. 41, 14, 2014.

[54] N. Gozlan. Poincaré inequalities and dimension free concentration of measure. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):708–739, 2010.

[55] B. Groux. Asymptotic Freeness for Rectangular Random Matrices and Large Deviations for Sample Covariance Matrices With Sub-Gaussian Tails. *arXiv.*

[56] A. Guionnet. *Large random matrices: lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[57] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000.

[58] A. Guionnet and O. Zeitouni. Large deviations asymptotics for spherical integrals. *J. Funct. Anal.*, 188(2):461–515, 2002.

[59] A. Hardy. A note on large deviations for 2D Coulomb gas with weakly confining potential. *Electron. Commun. Probab.*, 17:no. 19, 12, 2012.

[60] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2009.

[61] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.

[62] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[63] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, 1997.

[64] R. Killip and I. Nenciu. Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701, 2004.

[65] A. Klein. Extended states in the Anderson model on the Bethe lattice. *Adv. Math.*, 133(1):163–184, 1998.

[66] R. Latał a and J. O. Wojtaszczyk. On the infimum convolution inequality. *Studia Math.*, 189(2):147–187, 2008.

[67] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983.

[68] T. Leblé and S. Serfaty. Large Deviation Principle for Empirical Fields of Log and Riesz Gases. arXiv:1502.02970.

[69] M. Ledoux. A note on large deviations for Wiener chaos. In *Séminaire de Probabilités, XXIV, 1988/89*, volume 1426 of *Lecture Notes in Math.*, pages 1–14. Springer, Berlin, 1990.

[70] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[71] M. Ledoux. A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices. In *Séminaire de Probabilités XXXVII*, volume 1832 of *Lecture Notes in Math.*, pages 360–369. Springer, Berlin, 2003.

[72] M. Ledoux and B. Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15:no. 41, 1319–1343, 2010.

[73] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[74] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[75] J. B. Martin. Limiting shape for directed percolation models. *Ann. Probab.*, 32(4):2908–2937, 2004.

[76] P. Massart, G. Lugosi, and S. Boucheron. *Concentration Inequalities : A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[77] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.

[78] M. Maïda. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.*, 12:1131–1150, 2007.

[79] M. W. Meckes and S. J. Szarek. Concentration for noncommutative polynomials in random matrices. *Proc. Amer. Math. Soc.*, 140(5):1803–1813, 2012.

[80] S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.*, 7(5):745–789, 1979.

[81] A. Pizzo, D. Renfrew, and A. Soshnikov. Fluctuations of matrix entries of regular functions of Wigner matrices. *Journal of Statistical Physics*, 146(3):550–591, 2012.

[82] A. Pizzo, D. Renfrew, and A. Soshnikov. On finte rank deformations of Wigner matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(120):64–94, 2013.

[83] J. A. Ramírez, B. Rider, and B. Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.*, 24(4):919–944, 2011.

[84] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York, second edition, 1980. Functional analysis.

[85] R. T. Rockafellar. *Convex analysis.* Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997.

[86] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional integrals and derivatives.* Gordon and Breach Science Publishers, Yverdon, 1993. Theory and applications, Edited and with a foreword by S. M. Nikolʹskiĭ, Translated from the 1987 Russian original, Revised by the authors.

[87] B. Simon. *Orthogonal polynomials on the unit circle. Part 1*, volume 54 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, RI, 2005. Classical theory.

[88] B. Simon. Equilibrium measures and capacities in spectral theory. *Inverse Probl. Imaging*, 1(4):713–772, 2007.

[89] Ya. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)*, 29(1):1–24, 1998.

[90] Ya. G. Sinaĭ and A. Soshnikov. A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *Funktsional. Anal. i Prilozhen.*, 32(2):56–79, 96, 1998.

[91] R. Speicher. Free convolution and the random sum of matrices. *Publ. Res. Inst. Math. Sci.*, 29(5):731–744, 1993.

[92] R. P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

[93] M. Talagrand. A new isoperimetric inequality and the concentration of measure phenomenon. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 94–124. Springer, Berlin, 1991.

[94] M. Talagrand. The supremum of some canonical processes. *Amer. J. Math.*, 116(2):283–325, 1994.

[95] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.*, 6(3):587–600, 1996.

[96] R. C. Thompson. Convex and concave functions of singular values of matrix sums. *Pacific J. Math.*, 66(1):285–290, 1976.

[97] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

[98] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2003.

[99] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. Old and new.

[100] B. Virág. Operator limits of random matrices. In *Proceedings of the International Congress of Mathematicians, Vol. IV (Seoul, 2014)*, 2014.

[101] J. Von Neumann and E. Wigner. Uber das Verhalten von Eigenwerten bei adiabatischen Prozesses. *Phys. Z*, 30:467–470, 1929.

[102] V. H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.

[103] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[104] X. Zhan. *Matrix inequalities*, volume 1790 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2002.

Principes de grandes déviations pour des modèles de matrices aléatoires

Résumé : Cette thèse s'inscrit dans le domaine des matrices aléatoires et des techniques de grandes déviations.

On s'attachera dans un premier temps à donner des inégalités de déviations pour différentes fonctionnelles du spectre qui reflètent leurs comportement de grandes déviations, pour des matrices de Wigner vérifiant une propriété de concentration indexée par un paramètre $\alpha \in (0, 2]$.

Nous présenterons ensuite le principe de grandes déviations obtenu pour la plus grande valeur propre des matrices de Wigner sans queues Gaussiennes, dans la lignée du travail de Bordenave et Caputo, puis l'étude des grandes déviations des traces de matrices aléatoires que l'on aborde dans trois cas : le cas des β-ensembles, celui des matrices de Wigner Gaussiennes, et enfin des matrices de Wigner sans queues Gaussiennes. Le cas Gaussien a été l'occasion de revisiter la preuve de Borell et Ledoux des grandes déviations des chaos de Wiener, que l'on prolonge en proposant un énoncé général de grandes déviations qui nous permet donner une autre preuve des principes de grandes déviations des matrices de Wigner sans queues Gaussiennes.

Enfin, nous donnons une nouvelle preuve des grandes déviations de la mesure spectrale empirique des β-ensembles associés à un potentiel quadratique, qui ne repose que sur leur représentation tridiagonale.

Mots clés : Grandes déviations, matrices aléatoires, inégalités de concentration.

Large deviations problems for random matrices

Abstract: This thesis falls within the theory of random matrices and large deviations techniques. We mainly consider large deviations problems which involve a heavy-tail phenomenon.

In a first phase, we will focus on finding concentration inequalities for different spectral functionals which reflect their large deviations behavior, for random Hermitian matrices satisfying a concentration property indexed by some $\alpha \in (0, 2]$.

Then we will present the large deviations principle we obtained for the largest eigenvalue of Wigner matrices without Gaussian tails, in line with the work of Bordenave and Caputo. Another example of heavy-tail phenomenon is given by the large deviations of traces of random matrices which we investigate in three cases: the case of β-ensembles, of Gaussian Wigner matrices, and the case of Wigner matrices without Gaussian tails. The Gaussian case was the opportunity to revisit Borell and Ledoux's proof of the large deviations of Wiener chaoses, which we investigate further by proposing a general large deviations statement, allowing us to give another proof of the large deviations principles known for the Wigner matrices without Gaussian tail.

Finally, we give a new proof of the large deviations principles for the β-ensembles with a quadratic potential, which relies only on the tridiagonal representation of these models.

Key words : Large deviations, random matrices, concentration inequalities.